



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**The role of transparency in the  
acquisition of inflectional morphology:  
experimental studies testing  
exponence type using artificial  
language learning**

**Svenja Wagner**



**Submission in fulfilment of the degree of  
Doctor of Philosophy**

**School of Philosophy, Psychology and Language  
Sciences**

**The University of Edinburgh**

**2021**



## Acknowledgements

Writing this thesis has only been possible with the support and guidance of the people surrounding me over the past years.

I am grateful to my supervisor Jenny for letting me convince her of our shared linguistic interests at the first conference we met at, and for her continuous support ever since. Her piercing questions were of incredible help in reconsidering what we sometimes took for granted and to ensure we were asking the right questions. Being able to stop by her office for spontaneous feedback and exchange of thoughts made things so much easier. Whenever there were too many avenues to consider, she found a way to bring things back into perspective.

Kenny has been unmatched with his recollection of experimental techniques and his knowledge of programming technicalities. He also often pointed out those things in the bigger picture that would have otherwise been overlooked. I am grateful to him for his patience with my queries about the bits and pieces of statistics and for being an inspiration for work-life balance.

I am thankful to both my supervisors for their availability, care and guidance from start to finish, it has made all the difference.

Simon has left an imprint on this thesis with his intriguing questions and bringing in different perspectives during our regular supervision meetings. All three gave invaluable feedback in the many rounds of stimuli creation we had to endure. The support of my supervisors and the research group more generally made Edinburgh a very special place to be able to study at.

Thank you to all the associates, post docs, research assistants, visiting students and my fellow PhD students who were there during a shorter or a longer interval of my PhD for all the good conversations and opportunities for feedback over the years, the mutual support in experiment coding and all other sharing of experiment-related anecdotes. I happily remember the lunches in the Mosque Kitchen or under the Edinburgh sun with Andrés, Ash, Fausto, Fiona, Jonas, José, Tamar, Tom and many more.

Thank you also to my officemates, especially the two Ali's, Claire, Katerina, Soundess and Tom with whom I enjoyed many good moments at work in the quiet of evening hours.

I would like to thank Matt Spike for his recommendation of the paper by Igartua (2015).

I am also grateful to Linnea Stockall for her dedication during the workshop on morphological processing in Utrecht in 2019 and the engaging discussion we had as a part of it.

Deniz, Diana, Efe, Enes, Esperanza, José, Mora and Özge have shown incredible care during the creation of the Spanish and Turkish versions of my experiment.

Returning to the early phase, I would like to thank Alice Blumenthal-Dramé and Mirjam Eiswirth who supported my scientific interest prior to beginning my PhD and without whose guidance and encouragement my path would not have led me to Edinburgh in this form. Considering the final phase, I am grateful to Nina, for being a pillar of support during the quiet time of working from home.

I am grateful to everyone in my family who supported my research agenda, my late grandmum for always believing in me, my mum for every practical advice and her curiosity about my work, and my brother and dad for their reassurance especially during the last stage.

I am grateful to Abhinav for teaching me the concept of singular focus and being a role model for perseverance at work.

I would also like to thank the members of staff at the University of Edinburgh, especially Katie Keltie and Amanda Nisbet for being around and supportive with whichever administrative matter it was I needed to work out. Similarly, I am thankful to all the teaching staff who I was able to engage with over the years, especially Mits Ota and Jim Donaldson who I worked with the longest, and to Alisdair Tullo for his passion in solving any programming issues.

*“No explanation for linguistic phenomena is complete until a causal relation can be shown to exist between the principle proposed as explanation and the linguistic phenomena to be explained.” (Joan Bybee, 1985)*

## Summary

Languages can differ widely as to how they express grammatical meaning, such as time reference (as in present or past), whether one or more entities are being spoken about and who did what to whom. Some languages express these meanings as affixes to word stems, for instance the suffix *-ed* in *learned* indicates past tense. However, those languages that use affixes on word stems differ in whether a single affix expresses only one meaning or more than one. The suffix *-s* in *she learns* expresses three different meanings: third person, singular and present tense, while the *-s* in *learners* only expresses plural. Studies on the acquisition of languages in children and adults have proposed that those affixes that express a single meaning are easier to learn than those expressing multiple meanings, because the form-meaning mapping is more transparent. However, affixes in natural languages have many characteristics beyond transparency, for instance how regular or frequent they are. Since it is difficult to assess the impact of transparency on learnability in isolation in natural languages, I conducted a series of experiments using an artificial miniature language which allows me to test the effect of form-meaning mapping transparency of suffixes on learnability. The experiments were run with three groups of participants, who were native speakers of either English, Turkish or Spanish, in order to test whether the type of structures conveying grammatical meaning in a learner's mother tongue influence how well structures in a second language are learned. Across all experiments with over 600 participants, I do not find a learnability advantage based on form-meaning mapping transparency alone, which challenges previous claims about its importance for language learning. The level of transparency present in affixes in the mother tongue did not impact participants' learning in the experiment either. When looking at the types of errors made by participants, there is an indication of a weak bias for form-meaning mapping transparency, but taking the findings together, they suggest that learning ease is likely determined by transparency in combination with other factors like regularity and frequency. The thesis concludes by reconsidering some of the earlier language acquisition studies to highlight that the interplay of different factors was not always strictly accounted for, and suggests avenues for future research in this direction.

## **Abstract**

Agglutinating morphology has often been described as easier to learn than fusional morphology, in large part because it is more transparent (e.g., Brown, 1976; Goldschneider & DeKeyser, 2001; Igartua, 2015). Such claims mainly rest on data from natural languages, where transparency is often correlated with additional factors such as regularity, compositionality and frequency of morphemes in the system, making it difficult to quantify the effect of transparency alone. To address this issue, this thesis presents a series of artificial language learning experiments which make it possible to investigate the impact of transparency, and more specifically, separative exponence, on learnability in isolation. The artificial language instantiates a nominal paradigm with two binary features, CLASS and NUMBER, expressed through suffixes. In addition to controlling for factors like regularity, this paradigm size makes it possible to hold the number of morphemes to be learned constant. Transparency was manipulated between conditions by varying whether the morphemes exhibited one-to-one mappings between form and meaning (separative exponence, as in agglutinating systems) or one-to-many mappings (cumulative exponence, as in fusional systems). Experiments were conducted with three different groups of native speakers to represent different morphological techniques in the L1, namely native speakers of English (isolating), Turkish (agglutinating) and Spanish (fusional). This way, it was possible to test whether specific morphological experience from the L1 differentially influences the transparency benefit in the L2. Across all experiments, with over 600 participants in total, no learning advantage of the transparent condition over the non-transparent condition was found. Only a slight bias for form-meaning mapping transparency was revealed in participants' error patterns. These results suggest that the role of transparency in the acquisition of morphology is more limited than commonly assumed and that it may be better understood as a benefit only in combination with other characteristics of morphological paradigms such as regularity, generalisability and frequency of morphemes.

# Contents

<b>Preface</b> .....	11
<b>1. Introduction</b> .....	13
1.1 Morphological typology and the distinction between agglutinating and fusional types .....	14
1.2 Prototypical characteristics of agglutinating and fusional morphology .....	22
1.2.1 Exponence transparency .....	23
1.2.2 Polysemy and paradigm size .....	24
1.2.3 Regularity .....	27
1.2.4 System transparency .....	28
1.2.5 Word length .....	28
1.3 Difficulties in ascribing entire languages to a morphological type .....	31
1.3.1 Continuous development between morphological types .....	35
1.3.2 Geographical distribution .....	37
<b>2. Learnability differences across morphological types and the impact of L1 morphology</b> .....	41
2.1 The role of transparency for learnability .....	42
2.1.1 Transparency as the principle of 'one meaning one form' .....	43
2.1.2 Agglutinating structures are more transparent and more regular than fusional structures .....	46
2.1.3 Transparency as a proposed predictor of learnability in theory .....	47
2.1.4 The suggested role of transparency in studies of natural language acquisition .....	52
2.1.5 Transparency as a suggested driving force of language change .....	63
2.2 The role of compositionality and the generalisability of agglutinating systems .....	68
2.2.1 Definition of compositionality .....	68
2.2.2 Benefits of compositionality and generalisability during language learning .....	70
2.2.3 The development of compositionality in iterated learning studies in the lab and in computational models .....	72
2.3 Effects of the higher frequency of agglutinating morphemes .....	79
2.3.1 Agglutinating morphemes occur with a higher frequency than fusional morphemes .....	79
2.3.2 Implications of the higher frequency of morphemes for learnability .....	83
2.4 Summary of the effect of transparency and related predictors .....	84
2.5 The role of the mother tongue and transfer in the acquisition of L2 morphology .....	85
2.5.1 Native language transfer in L2 learning .....	86
2.5.2 Transfer of abstract categories from the L1 .....	86
2.5.3 Transfer effects arising from the morphological structure in the L1 ...	87
2.5.4 The differential processing of morphology in L1 speakers and L2 learners .....	88
2.6 Previous L2 research on implicit learning of artificial grammars .....	91
2.7 Research questions .....	95

<b>3. Testing the effect of transparency in artificial language learning in native speakers of English</b>	97
3.1 Preface	97
3.2 Paper: <b>The role of transparency in the acquisition of L2 morphology: artificial language learning experiments with native speakers of English</b>	99
3.3 Postscript	139
3.3.1 Choice of the categories NUMBER and CLASS	139
3.3.2 The relationship between CLASS and NUMBER in natural languages	141
3.3.3 Alternative options for testing NOUN CLASS systems	142
3.3.4 The use of categories with equipollent markers	145
3.3.5 Audio versions of the stimuli used in the experiments	148
3.3.6 Interpretability of the mean curves as an average across groups	148
3.3.7 Analysing learning rates using polynomial models	150
<b>4. Testing the effect of transparency in artificial language learning in native speakers of Turkish</b>	152
4.1 Introduction	152
4.1.1 Potential reasons for a lack of learnability difference in an English-speaking sample	153
4.1.2 The role of the native language in the acquisition of L2 grammatical morphemes	156
4.1.2.1 Transfer of grammatical categories and their morphological marking	158
4.1.2.2 Potential transfer of morphosyntax	160
4.1.2.3 Effects of L1 processing strategies on L2 morphological learning	163
4.1.3 Relevant aspects of Turkish morphology and grammar	170
4.1.3.1 Turkish as a synthetic agglutinating language	170
4.1.3.2 ANIMACY-based Differential Object Marking in Turkish	172
4.1.4 Predictions about learning outcomes for the L1 Turkish group considering language transfer	174
4.2 Methods	177
4.2.1 Participants	177
4.2.2 Creation of stimuli and additional experimental material	178
4.2.2.1 Artificial stems	178
4.2.2.2 Translations and creation of additional material	180
4.2.3 Procedure	181
4.3 Results	182
4.3.1 Overall accuracy	182
4.3.2 Discussion	184
4.3.3 Analysis of error patterns	184
4.3.4 Discussion	187
4.3.5 Analysis of questionnaire	188
4.3.5.1 Participants focusing on suffixes rather than stems	189
4.3.5.2 Participants reporting the correct classification system	189
4.3.5.3 Difficulties assessing meaning-morpheme mappings employed by participants	190
4.3.5.4 Participants who only learn one of the two features	190
4.3.5.5 Examples of different image classifications used	192

4.3.5.6	Participants focusing on single letters .....	193
4.3.5.7	Participants' awareness of the agglutinating nature of the artificial language .....	193
4.4	General Discussion .....	194
4.5	Conclusion .....	198
4.6	Appendix .....	199
<b>5.</b>	<b>Testing the effect of transparency in artificial language learning in native speakers of Spanish .....</b>	<b>201</b>
5.1	Introduction .....	201
5.1.1	Relevant aspects of Spanish morphology and grammar .....	202
5.1.1.1	Spanish as a synthetic fusional language .....	202
5.1.1.2	GENDER and ANIMACY distinctions in Spanish (and Basque) .....	203
5.1.2	Predictions about learning outcomes for the L1 Spanish group considering language transfer .....	206
5.2	Methods.....	208
5.2.1	Participants .....	208
5.2.2	Creation of stimuli and additional experimental material .....	209
5.2.2.1	Artificial stems .....	209
5.2.2.2	Translations and creation of additional material .....	211
5.2.3	Procedure .....	211
5.3	Results.....	212
5.3.1	Overall accuracy .....	212
5.3.2	Discussion .....	216
5.3.3	Analysis of error patterns .....	217
5.3.4	Discussion .....	221
5.4	Comparison of all L1 groups: English, Turkish and Spanish native speakers .....	222
5.4.1	Overall accuracy .....	223
5.4.2	Analysis of error patterns .....	225
5.4.3	Discussion .....	230
5.5	General Discussion .....	232
5.6	Conclusion .....	235
5.7	Appendix .....	237
<b>6.</b>	<b>General Discussion .....</b>	<b>240</b>
6.1	Aims of this thesis .....	240
6.2	Summary of predictions .....	242
6.3	Findings contradicting the claim surrounding the impact of transparency .....	246
6.4	Potential limitations of the experimental design .....	248
6.5	Implications of the experimental results .....	249
6.6	The problematic focus on transparency in acquisition research based on linguistic theory .....	252
6.7	Why paradigm size matters .....	260
6.8	Summary and outlook .....	263
<b>7.</b>	<b>Conclusion .....</b>	<b>266</b>
<b>8.</b>	<b>References .....</b>	<b>268</b>
<b>9.</b>	<b>Appendix .....</b>	<b>279</b>

## Preface

The languages of the world differ in some remarkable ways. Some convey grammatical meaning largely without inflections, such as Vietnamese and English. Others are inflectionally rich, like Russian, Swahili and Turkish. Among the latter, one striking difference concerns the way languages express grammatical meaning through overt forms and how they compose words. The relationship between grammatical forms and the meanings they convey has been investigated over many decades, one frequently asked question being in what way this form-meaning relationship impacts language acquisition, both in a first language (L1) and in a second language (L2). Transparency is often invoked as a notion to describe the obviousness of the link between a form and its meaning and it has been claimed that the more transparent the link is, the earlier the form will be acquired. Transparency levels differ greatly between two prototypical types of inflectional morphology, namely agglutinating and fusional systems. Therefore, claims about the proposed effect have sometimes been made against this backdrop. This thesis has its focus on the effect of transparency and the broader context of these two morphological types, investigating the impact of the transparency of the systems' morphemes, which can be described as "the smallest meaning-bearing unit[s] of language" and the building blocks of words (Kortmann, 2005). While agglutinating and fusional structures differ in many additional ways in natural languages, I single out the transparency of exponents to test its proposed significance for language acquisition. To do so, this thesis addresses the question of learnability through an experimental research paradigm in which I manipulate the level of exponence transparency in an artificial language learning task<sup>1</sup>.

The context of this thesis is defined by properties of morphological systems whose classification goes back to the pioneers of morphological typology. In **Chapter 1**, a brief excursion into the development of the field and the origins of some of the terminology used throughout this thesis lays the groundwork for the chapters to follow. I then look at the ways in which agglutinating and fusional systems differ more generally, since morphemes are typically part of larger inflectional paradigms. Having noted the broader context, in **Chapter 2** I lay out the claims that have been made with

---

<sup>1</sup> Since the research context in which this thesis is situated is concerned with transparency in a broad sense, I use this term throughout. However, the experimental studies I conducted strictly only test the transparency of exponents, which contributes to the more general notion of transparency.

respect to transparency more specifically, and how it is predicted to impact learning ease. I illustrate observations from first and second language acquisition literature as well as studies on language change which trace acquisition ease, at least in part, back to the transparency of morphemes. However, due to the nature of morphological paradigms in natural languages, it is difficult to separate transparency from other factors such as the regularity of inflectional systems, their compositionality and generalisability, as well as the frequency of morphemes. Indeed, some of the earlier studies on the role of transparency in language acquisition remain vague as to which aspect of transparency they refer to. I therefore devised an artificial language learning paradigm that allows me to do exactly this, namely isolate one aspect of transparency, exponence transparency, and test its effect in controlled inflectional systems.

Based on existing research on transfer effects of morphological structures from the L1 onto an L2, the question arises whether experience with morphological techniques in the mother tongue influences the acquisition of morphology in an L2. **Chapters 3-5** each describe experiments targeted at one particular L1 speaker population, namely adult native speakers of English, Turkish and Spanish. These languages were chosen because they can be largely attributed to isolating, agglutinating and fusional types, respectively, and therefore enable me to test the role of the morphological type and the morphological richness of the mother tongue in the acquisition of second languages. In each experiment, the success of learning is being compared between an agglutinating and a fusional system, with the main difference between them being the level of transparency in the form-meaning mapping of their morphemes. **Chapter 3**, relating the results of the L1 English participant group, is based on a paper under review, which will be contextualised in the thesis. **Chapter 4** describes the findings from an experiment with native speakers of Turkish and **Chapter 5** describes the findings from an experiment with Spanish native speakers. Generally speaking, I find no significant effect of the transparency of form-meaning mappings on acquisition across all three groups. In the **General Discussion** in **Chapter 6**, I bring the findings from the individual chapters of this thesis together with the observation that transparency stands in close relation with other factors in inflectional systems and conclude that the impact of transparency alone appears to be overstated. I illustrate the difficulties of investigating it separately in naturally occurring languages on the basis of a selection of published studies that make claims about transparency leading to acquisition ease. I conclude by highlighting the contribution of

this thesis to research on the role of transparency in language acquisition. I will end by issuing a word of caution in favour of more rigorous tests of the causal relationship between transparency and learnability and by outlining potential avenues of research that could shed additional light on the interplay between transparency and other factors.

## **Chapter 1**

### **Introduction**

First, however, the introduction in **Chapters 1** and **2** will illustrate areas of research that have investigated the acquisition of inflectional morphology. **Chapter 1** commences with an overview of the beginnings of the field of morphological typology, followed by a discussion of the morpheme, which is a central notion and unit of analysis in this thesis. I then describe how agglutinating and fusional morphology are cover terms for a range of different characteristics, transparency between forms and their meanings being a central one of them. **Chapter 2** discusses characteristics of morphological systems that likely have an impact on language acquisition. In **Section 2.1** I provide a definition of transparency before summarising different claims that have suggested a link and a causal effect between transparency and acquisition ease. This sets the scene for the subsequent sections, in which I will outline those factors that likely also play a role in the acquisition of grammatical morphemes, since they are often confounded with transparency in natural languages: **Section 2.2** discusses compositionality and generalisability and **Section 2.3** discusses frequency as factors that are closely related to the transparency of morphemes in inflectional systems, but which are not discussed in depth in studies reporting on the impact of transparency on natural language acquisition. One of the messages throughout **Chapters 1** and **2** and the thesis more generally is that such a correlation between characteristics makes it difficult to understand the contribution of individual factors to learning ease in natural languages. I will return to the role of these four factors in **Chapters 3** and **6**. **Section 2.5** presents an overview of transfer effects between L1 and L2 in general and in relation to the acquisition of morphology in particular. The question whether knowledge of morphological techniques from the L1 affects the acquisition of morphology in an L2 is addressed in **Chapters 3-5** and summarised in **Chapter 6**.

**Chapter 2** concludes with a review of studies testing the implicit learning of grammatical features, which links to the methodology employed in the experiments presented in **Chapters 3-5**.

## **1.1 Morphological typology and the distinction between agglutinating and fusional types**

When considering the kinds of morphology found across languages, certain patterns in their variation can be established. While some languages largely lack inflections, others have different types of affixes and ways of modifying word stems, and these differences are not random. At the same time, it is generally agreed that inflectional morphology presents a difficult area of language for L2 learners to master (Clahsen et al., 2010; DeKeyser, 2005; Prévost & White, 2000). This has sparked an interest in determining factors that lead to this difficulty, and it is necessarily connected to the kinds of morphology being acquired. For this reason, I begin with a brief review of the development of morphological typology as a research field and the notion developed in this field which is most relevant to this thesis, namely morphological types. Although this conceptualisation is idealistic (Comrie, 1981; Dressler, 2003; Dressler et al., 2006; Haspelmath, 2009), it will be shown that subsequent insights from the early attempt to classify whole languages based on their morphological systems can help us understand what types of structures and what combinations of factors might cause the observed acquisition difficulty.

A central aspect of this investigation is the notion of the morpheme, which was already invoked as a unit of analysis in the theoretical description of languages by Jan Baudouin de Courtenay in the 19<sup>th</sup> century, used by Structuralist scholars like Leonard Bloomfield and Zellig Harris in the 20<sup>th</sup> century and remains in use until the present day (Anderson, 2015; Stewart, 2018). Baudouin de Courtenay regarded the morpheme as a part of a word which has psychological autonomy and cannot be further subdivided, which was taken up in later research (Anderson, 2015). Generally in line with other Structuralists, Bloomfield (1933) defines the morpheme as “a linguistic form which bears no partial phonetic-semantic resemblance to any other form”, by which he means that the morpheme cannot be further subdivided into a smaller unit of a form-meaning relationship which also occurs in other forms (Bloomfield, 1933, p.161, as quoted in Anderson, 2015, p.14). According to Bloomfield

(1933), any complex form is constituted of morphemes, a conceptualisation which relates to the present-day use of the term ‘morpheme’ as a smallest unit, although Bloomfield (1933) appears to view the morpheme as having a concrete phonological form (Stewart, 2018). This was developed further by other Structuralist scholars who regarded the morpheme as an abstract element, like the phoneme, which was instantiated by concrete elements, namely allomorphs (Anderson, 2015). Similarly, Harris’ method to obtain a description of the grammatical structure of a language consists of breaking up any utterance into elements (phonemes and morphemes) and to subsequently describe the distributional relationships between them. Thereby, elements are defined as “the minimum, i.e. smallest distributionally independent, descriptive factors (or elements)” of an utterance (Harris, 1951, p.21). Today, the conceptualisation of the term ‘morpheme’ differs between morphological theories, but for the practical purpose of describing relationships between forms and meanings as pursued in this thesis, we can still understand it as the smallest abstract unit of meaning that utterances or words can be segmented into (Anderson, 2015; Kortmann, 2005). I will illustrate this central unit of observation with some examples before moving on to the development of the field of morphological typology.

Morphemes as units of meaning can coincide with words, but do not have to necessarily. For instance, the word *learner* consists of two morphemes, *learn* and *-er*. The first is the stem indicating an action and the second indicates a DOer, a person who learns. By contrast, the word *barrier* cannot be segmented any further, because we cannot find any smaller units that carry independent meaning. A further technical distinction can be made between lexical meaning such as *-er* in *learner* and grammatical meaning such as *-s* expressing plural as in *learners*. Morphemes are abstract units and stand for the meaning to be conveyed, such as the plural morpheme. However, a morpheme can have more than one instantiation, for example depending on the lexeme in question (Greenberg, 1960; Kortmann, 2005). While morphemes which do not correspond to neatly segmentable affixes posed difficulties to Bloomfield’s (1933) analysis, Harris broadened the scope of morpheme variation to include also non-segmentable and zero morphemes (Harris, 1951; Stewart, 2018). The case of the German plural illustrates the variety of morphemic alternations because German has various formal means of expressing plural meaning. Which marker is chosen depends on the GENDER and the phonological properties of the lexeme in question (Szagun, 2001). To give some examples, plural can, among others,

be indicated through the plural suffix *-e* as in *Laut – Laute* (sound – sounds), *-en* as in *Frau – Frauen* (woman – women), or *-n* as in *Zeile – Zeilen* (line – lines), where the plural form is added to the end of the stem. The endings *-e*, *-en*, *-n* are examples of different morphs (concrete forms) instantiating the plural morpheme (see Kortmann, 2005). Morphs like *-e*, *-en*, *-n* are the actual form units contained in concrete utterances, and all morphs expressing the same meaning belong to the same morpheme, which is the corresponding abstract meaning unit (Greenberg, 1960). While there is a technical difference between morphs and morphemes, these differences are sometimes conflated and both are referred to as morphemes. Throughout this thesis, I will use this simplification unless the relation between the two distinct notions needs to be pointed out to illustrate an argument. Further, not in all cases are individual morphs linearly aligned in utterances, as we will see throughout this thesis. An example of this is the German 2<sup>nd</sup> person singular present tense verb *spielt* ('you play'), where the ending *-t* indicates all grammatical meanings at once. There has been great interest in classifying languages according to the morphological techniques they display, which draw on the relationship between morphemes, their instantiations through morphs, and their meanings<sup>2</sup>.

Research on morphological techniques has a long history. The field of morphological typology goes back to the first half of the 19<sup>th</sup> century and was dominant during the 19<sup>th</sup> and early 20<sup>th</sup> century (Blumenthal-Dramé & Kortmann, 2013; Bybee, 1985; Comrie, 1981). Some of its classifications are important until the present day, although the distinctions made and the understanding of their absoluteness have changed over time, as will be shown below. Morphological typology saw its beginning when Friedrich von Schlegel (1772-1829) classified languages into those with affixes and those with inflections. According to Greenberg (1960), Schlegel described affixes as mechanically put together, and saw them as inferior to the inflectional type, which coincided with the structures found in many Indo-European languages. The former

---

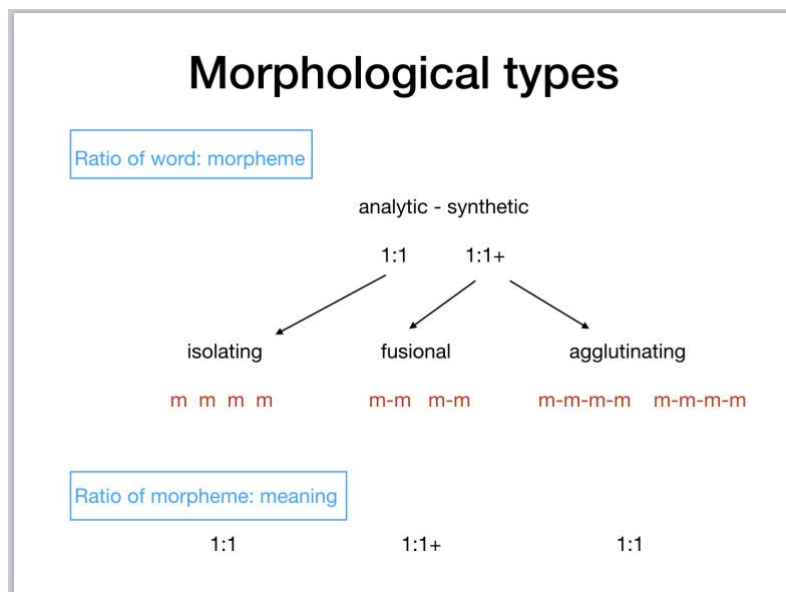
<sup>2</sup> The term *morphological technique* might suggest an agency of the language to some readers, which is not my intention. As I will show in **Section 1.3**, I am avoiding the terms *morphological type* and *language type* since they suggest that a language consistently displays one kind of morphology. Similarly, *morphological system* appears too broad a term when describing the morphological make-up of a specific part, for instance noun morphology, in a language. Therefore, I speak of *morphological technique* to highlight the possibility that this may refer to a small sub-part of the grammar of a language, such as noun morphology or even individual inflectional classes. The term *technique* with respect to morphology is not uncommon either and therefore allows for consistency of terminology (see, e.g., Bickel & Nichols, 2009; Greenberg, 1960; Haspelmath et al., 2008 [Chapter VIII: Morphological techniques]).

type is what is today known as agglutinating. Such evaluative attitudes prevailed for a while, but it is remarkable how agglutinating-type structures later came to be regarded as more optimal, desirable, evidently more common and therefore more default-like than fusional structures, as we will see in **Section 2.1** (Bickel & Nichols, 2013b; Bybee, 1985; Carstairs-McCarthy, 1987; Don, 2017). August Wilhelm von Schlegel (1767-1845) added a third type, namely languages without grammatical affixes, the isolating type. He also introduced the idea of analytic and synthetic languages, but as sub-types of the inflectional type, which differs from their present-day distinction (Haspelmath & Michaelis, 2017). Wilhelm von Humboldt (1767-1835) later added the incorporating type to the classification system. His classification into isolating, agglutinating, fusional and incorporating was taken up by August Schleicher (1821-1868) and Friedrich Max Müller (1823-1900) and resembles that of the present day, although the emphasis is often on the first three types (Comrie, 1981; Haspelmath, 2009).

However, criticism of the early valuating stance in favour of fusional types taken up by some European linguists was building up, voiced, for instance by Fritz Mauthner (1849-1923), who called the judgmental ranking of whole languages according to their level of transparency of inflections “foolish” (Greenberg, 1960, p.181). Sapir (1921) takes a similar non-evaluative stance and writes that “[a] linguist that insists on talking about the Latin type of morphology as though it were necessarily the high-water mark of linguistic development is like the zoölogist that sees in the organic world a huge conspiracy to evolve the race-horse or the Jersey cow” (p.131). He also added a further basis for distinction, namely between analytic, synthetic and polysynthetic types, reflecting the number of morphemes that are concatenated into words (Greenberg, 1960; Sapir, 1921). Today, three morphological types are mainly distinguished – isolating, agglutinating and fusional types. Some scholars also include two additional ones – polysynthetic and incorporating – which are co-exhibited in many languages (Comrie, 1981; Penke, 2012). The term polysynthesis was coined by Peter Duponceau in 1816 and generally speaking refers to the combination of a considerable number of morphemes, which can be lexical as well as grammatical (Comrie, 1981; Fortescue et al., 2017). The incorporating type was first described by Wilhelm von Humboldt in the late 1820s and is characterised by the combination of two lexical items into a complex word form, rather than a lexical and a grammatical form (Bybee, 1985; Fortescue et al., 2017). This makes the incorporating type a sub-type of the polysynthetic one (Comrie, 1981). As Bybee (1985) puts it, incorporation typically

“refers to the fusion of the nominal patient of the verb with the verb, but often two verb stems can be fused as well” (p.105). Comrie (1981) also gives the example of the English compound swimsuit (p.42). While the more general term of polysynthesis is taken to describe a very high level of morphological complexity due to the number of morphological elements within a word, there is ongoing debate as to the core features of this type (Fortescue et al., 2017). It is further noteworthy that early morphological typology tried to assign entire languages to morphological types, a strategy that is still sometimes found today, despite considerable criticism of such a holistic approach (e.g., Greenberg, 1960; Haspelmath, 2009; Sapir, 1921). I will address this observation in more detail below.

**Figure 1** illustrates how the different morphological techniques stand in relation to one another according to common typological classification (Blumenthal-Dramé & Kortmann, 2013; Comrie, 1981; Haspelmath & Michaelis, 2017). The differentiation made here mainly rests upon the ways in which (subsystems of) languages map grammatical meaning onto morphemes and how morphemes are grouped together to form words. Note that **Figure 1** illustrates the three types that are most relevant for this thesis. The analytic-synthetic layer describes languages which use mainly free morphemes, where morphemes typically do not group together to form words, (analytic) on one end and languages where words typically consist of more than one morpheme (synthetic) on the other. The second layer is concerned with the relation between individual morphemes and the meanings they express. The isolating and the agglutinating types are examples of analytic and synthetic languages, respectively, and they both typically convey each grammatical meaning by a separate morpheme, which means that their form-meaning mapping is transparent (Blumenthal-Dramé & Kortmann, 2013; Comrie, 1981). **Figure 1** illustrates how this technique can lead to relatively long words in agglutinating systems. Morphemes of fusional types express multiple meanings at once, leading to shorter words in comparison to agglutinating words (see Plungian, 2008). This difference in word length may have implications for learnability and will be addressed in **Chapter 3**.



**Figure 1:** Overview of some morphological types ('m' denotes morpheme, hyphens indicate within-word boundaries, spaces indicate between-word boundaries)

In this regard, the example of the German verb *spiel-t* ('you play') from before can be contrasted with the Turkish verb *konuşu-yor-sun-uz* ('you are speaking'). They both express the same number of grammatical features (PERSON, TENSE, NUMBER), but require a different number of inflectional morphemes to do so. *Spielt* indicates present tense, 2<sup>nd</sup> person and plural through the single morpheme *-t* while *konuşuyorsunuz* expresses these meanings through separate morphemes: *-yor* (present tense continuous), *-sun* (2<sup>nd</sup> person), and *-uz* (plural). While relatively prototypical examples among natural languages have been identified, such as Turkish (agglutinating), Latin (fusional), Annamite (root isolating), Chukchi (incorporating) and Eskimo (polysynthetic), **Figure 1** is clearly a simplification since morphological types often display additional characteristics, as I will describe below (Comrie, 1981; Greenberg, 1960; Haspelmath, 2009; Plungian, 2008). Nevertheless, for means of illustration, the examples in (1) and (2) show a nominal paradigm in Turkish and in Russian, which are regarded as rather prototypical examples of the agglutinating and the fusional type, respectively.

(1) Declension of the Turkish noun *adam* ('man')

	Singular	Plural	
Nominative	<i>adam</i>	<i>adam-lar</i>	
Accusative	<i>adam-ı</i>	<i>adam-lar-ı</i>	
Genitive	<i>adam-ın</i>	<i>adam-lar-ın</i>	
Dative	<i>adam-a</i>	<i>adam-lar-a</i>	
Locative	<i>adam-da</i>	<i>adam-lar-da</i>	
Ablative	<i>adam-dan</i>	<i>adam-lar-dan</i>	(adapted from Comrie, 1981, p.41)

(2) Declension of the Russian noun *stol* ('table')

	Singular	Plural	
Nominative	<i>stol</i>	<i>stol-y</i>	
Accusative	<i>stol</i>	<i>stol-y</i>	
Genitive	<i>stol-a</i>	<i>stol-ov</i>	
Dative	<i>stol-u</i>	<i>stol-am</i>	
Instrumental	<i>stol-om</i>	<i>stol-ami</i>	
Locative	<i>stol-e</i>	<i>stol-ax</i>	(adapted from Comrie 1981, p.41)

Despite its simplifications, **Figure 1** illustrates features that play an important role in the experiments reported in this thesis. The transparency between morphemes and their meanings is the central criterion I manipulate and the morphological systems I test reflect the tendency of words to differ in length depending on morphological type. This basis of classification is similar to what is found in some of the literature. For example, Plungian (2008) states that “[t]he terms “agglutination” and “(in)flexion”, taken in their most general and most established sense, oppose two major techniques of morpheme combination within a word-form, or, to put it slightly differently, two ways of assembling words in natural languages” (p.669). Further below, I will show to what degree within-word morpheme boundaries are respected depending on morphological type (**Section 1.2.1**) and I will list additional characteristics that typically co-occur with agglutinating and fusional types, since these two are the focus of the present analysis.

First, however, it is noteworthy that several terms exist for morphological types. This thesis uses *agglutinating* as a synonym for *agglutinative* and *fusional* for what is also referred to as *flexive*, *flectional* or even *inflecting* by some (see, for instance, Laaha & Gillis, 2007; Plank, 1986; Plungian, 2008). For example, Laaha & Gillis (2007) label as ‘inflecting’ what I label ‘fusional’ in this thesis. Crucially, I regard both

agglutinating and fusional morphology as subtypes of inflectional morphology in that both types are part of the synthetic type which uses inflections on word stems in order to mark grammatical meaning. This is in line with established and recent work by other scholars (Comrie, 1981; Kortmann, 2005; Haspelmath, 2009). The synthetic types stand in contrast to analytic types, which do not use much morphology to express grammatical meaning. I adopt the terms ‘agglutinating’ and ‘fusional’ because they appear most widely used at present (Haspelmath, 2009) and because the labels speak relatively directly to the type of exponence used, which is a central criterion as outlined below. As noted, further types have been labelled by scholars, but a more comprehensive overview is beyond the scope and focus of this thesis.

As mentioned above, the classification of morphological types was initially applied to whole languages and notably, this conceptualisation is still found in present-day linguistics textbooks (Kortmann, 2005; see also Haspelmath, 2009). While such simplifications are arguably beneficial to beginning students of linguistics learning these concepts, the classification may become less helpful when it forms the basis of typological studies. Indeed, the use of the terms ‘agglutinating’ and ‘fusional’ has been under some rightful criticism (Comrie, 1981; Haspelmath, 2009; Sapir, 1921). Two main areas of concern will be addressed here. Firstly, the concepts of ‘agglutinating’ and ‘fusional’ carry many characteristics under the hood, which makes it difficult to define what exactly the terms are supposed to mean (see Haspelmath, 2009).

A second problem is that speaking of agglutinating and fusional *languages* may imply that a language’s morphology consistently follows one type or the other. A statement from Hengeveld & Leufkens (2018) exemplifies this. In a cross-linguistic investigation of the transparency of languages, they describe their selection in the following way: “[T]he languages in the sample are of various types. We took special care to include fusional, agglutinative, and isolating languages, as well as polysynthetic ones, as the morphological type of a language seems to correlate to some degree with its transparency” (p.159). Their choice is sanctioned since they also use an aggregate measure of transparency and look for more general correlations. However, the picture in natural languages is often more complex and simplifications may need to remain restricted to appropriate contexts. I will consider both of these issues in turn.

## 1.2 Prototypical characteristics of agglutinating and fusional morphology

Regarding exact definitions of both types, it is important to take account of the fact that ‘agglutinating’ and ‘fusional’ are cover terms for a range of characteristics that have been described in connection with these types, although languages may display them to varying degrees. Another related observation is that for example, Haspelmath (2009) makes a distinction between characteristics that are implicit in the notion of agglutinating and fusional types on the one hand, and correlations between these types and other features on the other hand. This is because he is putting the term ‘agglutinating type’ to the test. My focus, however, is different: I am interested in the contribution of morpheme transparency to a system’s learnability, and it is important to take co-occurrences with other characteristics into account to understand whether a combination of factors might be at play. However, a lengthy discussion of all these factors is only helpful when they are also tested. Therefore, I will provide a non-exhaustive list of commonly stated characteristics, highlighting especially those that are of relevance to the research questions pursued in this thesis. Most of the characteristics shown in **Table 1** are discussed in the present sub-section. Since the impact of compositional structure and the frequency of morphemes on learnability have been tested in a multitude of studies which are important for the contextualisation of the findings of this thesis, these characteristics are discussed separately in **Sections 2.2** and **2.3**. While it is important to consider the interplay of features that we find in agglutinating and fusional structures, this thesis tests one specific aspect and its role in language acquisition, namely exponence type.

<b>Characteristic</b>	<b>Instantiation typical of agglutinating structures</b>	<b>Instantiation typical of fusional structures</b>
Exponence transparency	separative exponence	cumulative exponence
Variance of affixes and stems	invariance (lack of allomorphy)	variance (allomorphy)
Inflectional classes	lack of inflectional classes	inflectional classes
Syncretism	lack of syncretism	syncretism
Paradigm size	larger paradigms	smaller paradigms
Regularity	more regularity	less regularity
System transparency	transparency	opacity
Word length	longer words	shorter words
Compositionality of morphemes	compositional structure	holistic structure
Frequency of individual morphemes	higher frequency	lower frequency

**Table 1:** Typical characteristics of agglutinating and fusional structures

### 1.2.1 Exponence transparency

A central distinguishing feature often referred to in the literature is that of exponence, which is also the main criterion investigated in this thesis (Bickel & Nichols, 2013b; Bybee, 1985; Igartua, 2015; Plungian, 2008). Exponence relates to the way grammatical meaning is mapped onto morphemes: agglutination is characterised by separative exponence of grammatical meanings – each morpheme typically only carries one meaning. Fusion, on the other hand, is characterised by cumulative exponence – a single grammatical morpheme expresses more than one grammatical meaning, but cannot be further segmented into meaningful elements. The labels are telling, with agglutination going back to Latin *agglutinare* (‘fasten with glue’), which captures the idea that agglutinating morphemes are glued together and invariant in their shape (Online Etymological Dictionary). In the case of fusion, on the other hand, two or more meanings are, quite literally, fused into a single form. Agglutinating morphemes are typically combined linearly to form words and multiple cumulative formatives can also be combined linearly in a word form (Bickel & Nichols, 2013b;

Igartua, 2015; Plank, 1986; Plungian, 2008). Plungian (2008) refers to separative exponence as “the main semantic parameter of agglutination” (p.672). Recall the example of the fusional German verb *spiel-t* and the agglutinating Turkish verb *konuşu-yor-sun-uz*, which both express TENSE, PERSON and NUMBER, but the mapping of these inflectional categories onto morphemes differs. In German, they are all mapped onto a single morpheme, while separate morphemes are used in Turkish. *Konuşuyorsun*, by means of comparison, means ‘you speak’ in the singular, since the suffix *-uz* in the plural is missing. The role of this transparency and corresponding compositionality (or lack thereof) and its implications for a learner are central to this thesis, and will be discussed further in **Sections 2.1** and **2.2**.

### **1.2.2 Polysemy and paradigm size**

In addition to exponence, further characteristics of agglutinating and fusional types have been proposed in the literature. Agglutinating systems typically display affix and stem invariance, that is a lack of allomorphy and other instances where the form of the morpheme varies (Haspelmath, 2009). Fusional systems, on the other hand, are characterised by allomorphic variation (also referred to as alternation) of stems and affixes (Haspelmath, 2009; Plungian, 2008). Haspelmath (2009) interprets cumulation and variance as that “in a typical agglutinating language, each affix not only stands for just one subcategory, but is also invariant in its shape, whereas in fusional languages, not only stems, but also affixes show considerable morphophonological allomorphy” (p.17). Notably, some scholars treat cumulation and stem and affix alternation as two different instances of fusion. For example, Hengeveld & Leufkens (2018) see both as a case where “the boundaries between individual morphemes within a word are not respected” (p.149). Crucially, I treat both as separate features, although they are arguably related: cumulation is a more extreme case of the neglect of morpheme boundaries than stem and affix alternation. However, since I treat fusion as the cover term of a morphological type, it necessarily encompasses additional features according to my classification. Reducing fusion to these two instances would result in using the term for two separate concepts. Instead, I use the term cumulative exponence when speaking of the fusion of morphemes in a single form. Several additional characteristics of morphological types are outlined further below.

Moreover, Plungian (2008) differentiates between phonologically conditioned allomorphic variation and grammatically conditioned variation. In the former, the shape of the morpheme varies depending on the phonological environment, while in the latter, it depends “on semantic or grammatical properties of an adjacent morpheme” (Plungian, 2008, p.670, see also Kusters, 2003). An example of phonological conditioning is the English plural form -s, which can take the allomorphs /-z/, /-s/ and /-ɪz/ depending on the preceding sound. In cases of lexical conditioning, individual lexemes require a certain allomorph. Grammatical conditioning, also referred to as morphological conditioning, is at play when a grammatical morpheme requires a particular form of another morpheme (usually the stem) (Kortmann, 2005, pp.90-92; Plungian, 2008). An example is the verb *sleep*, which combines with the morpheme *-ing* without modification, but shows a change in stem in its past tense form *slept* (Kortmann, 2005, p.92). The fact that these processes are paralleled in other lexemes (see *weep* – *wept*) motivates Kortmann’s (2005) analysis of them as grammatically, and not lexically, conditioned. Within grammatical conditioning, there are extreme cases where an allomorph not only shows alternation, but shows no resemblance to the root morpheme. In these instances, we speak of suppletion, as in the English forms *good* – *better* and *go* – *went* and the Latin suffixes *-o* and *-m* for 1<sup>st</sup> person singular subjects in the present tense and in the imperfect tense, respectively (Haspelmath, 2009, p.18; Kortmann, 2005, p.92). Hengeveld & Leufkens (2018) categorise irregular stem formation and suppletion as examples of morphological conditioning. They even take a slightly more radical stance in claiming that stem and affix alternation are tightly connected to morphological type, in that the agglutinating type displays it to a limited extent, while the fusional type by definition shows this non-transparent feature. In short, morpheme variance adds a layer of irregularity and additional forms to fusional systems.

A related distinctive feature of fusional types is the existence of conjugation and declension classes, for instance as they occur in Latin (Plungian, 2008). Plungian (2008) views these as instances of lexically conditioned affix variation where different allomorphs are chosen depending on the lexical stem. Agglutinating systems, by contrast, are characterised by uniformity of both stems and affixes (Haspelmath, 2009). In other words, an agglutinating morpheme tends to be expressed by a single morph, which does not change depending on context. The main insight to take away from these illustrations is that variation adds further forms to a morphological

paradigm, which might be described as additional complexity in a general sense (Karlsson et al., 2008; McWhorter, 2007; Sims-Williams & Enger, 2020). Moreover, Kortmann (2005) speaks of grammatically conditioned variation as irregularity. However, also the seemingly more regular phonological variation in a way adds irregularity to a morphological paradigm, since different forms exist to express one morpheme, such as the English plural. Crucially, these types of variation have been found predominantly in fusional systems, while agglutinating morphology is characterised by higher regularity. I will return to this link between morphological type and regularity in **Section 2.1**.

Fusional systems have also been associated with syncretisms between inflectional markers and with smaller paradigms (Plank, 1986). Agglutinating systems, by contrast, display a lack of polysemous forms and have been linked to larger paradigms (Plank, 1986; Plungian, 2008). The suggested relationship between these features and exponence type is instructive to consider in light of the research questions of this thesis. Since polysemy reduces the transparency of forms, it is linked to claims about the learnability of morphemes of different levels of transparency. Syncretic morphs express different meanings (morphemes) depending on the context in which they occur (Kortmann, 2005). One example is syncretism within the same declension class. Plungian (2008) refers to the cumulative NUMBER/CASE declension of the Russian noun *stena* (wall), where the genitive singular form *sten-ý* only differs from the nominative plural form *stén-y* in terms of stress placement (p.672). The German verb *spielen* also expresses 1<sup>st</sup> or 3<sup>rd</sup> person in the plural depending on the context and examples of German syncretic plurals such as *Schatten – Schatten* (shade – shades) show no formal distinction between singular and plural.

The richness of inflectional systems in syncretic forms may be intimately linked to the size of the paradigm: As Plank (1986) illustrates, the amount of affix variation that correlates with fusional paradigms can quickly reach a large scope. Additionally, fusional techniques require more morphemes than agglutinating techniques to express the same number of complex meanings, which is a consequence of exponence type. I will expand on this observation in **Section 2.3**. In order to retain a system which can express all the intended grammatical meanings, but is learnable at the same time, fusional languages may develop syncretic morphs. While syncretisms might therefore be a practical consequence of other features of fusional systems, namely affix variation and cumulative exponence, it adds to the opacity of the form-

meaning relationship. What agglutinating systems save – in terms of number of formatives – (with their affix invariance, affix uniformity and lack of inflectional classes) can be used for the number of values that a grammatical feature can take. For instance, if a language has few CASE allomorphs, a learner may be capable of acquiring more CASE distinctions instead. This might explain why agglutinating systems resort less to syncretisms and why agglutinating paradigms can be larger than fusional paradigms at the extreme end, as has been claimed in the literature: they have more capacity to express more grammatical meanings through fewer forms (Plank, 1986). This observation is important – it links to research on the advantages of compositional systems that will be introduced in **Section 2.2**. Paradigm size will also play a role for the arguments presented in this thesis. I will return to it in **Chapters 3 and 6** when summarising the experimental findings and delineating the potential impact of transparency on acquisition.

### **1.2.3 Regularity**

As a product of lack of inflectional classes as well as affix alternation, lack of exceptions and consistent overt expression, typical agglutinating systems are also more regular than fusional systems, as has been observed many times. In multiple instances, Greenberg (1960) relates agglutination to regularity, partly implicitly. However, he also explicitly links agglutination to morphological regularity. Plungian (2008) describes agglutinating morphemes as displaying uniformity and regularity. Haspelmath (2009) also observes that “[t]he basic idea is that in a typical agglutinating language, each affix not only stands for just one subcategory, but is also invariant in its shape, whereas in fusional languages, not only stems, but also affixes show considerable morphophonological allomorphy” (p.17). When a grammatical meaning is only expressed through a single form and this form only has a single meaning in the language, this is a case of biuniqueness. An example is the English affix *-st* to express superlative meaning (Dressler, 2012). If affixes tend to be invariant, this reduces the number of different forms and also the number of subclasses in inflectional paradigms. In addition, it renders the concrete forms of morphemes (i.e. the morphs) more predictable. Compared to this, a fusional system requires more forms and the conditions under which these forms are to be used can make the system more

complex. Especially in cases of lexical conditioning, this adds a layer of irregularity to fusional systems.

#### 1.2.4 System transparency

The discussion of typical characteristics of agglutinating and fusional systems has shown how many of them may be linked. Additionally, a consequence of exponence type, affix and stem variance, inflectional classes, syncretism and regularity is the resulting transparency, or lack of transparency, of morphemes, words and paradigms more broadly. This can be summarised as system transparency. For instance, separative formatives are more transparent than cumulative ones because the link between a form and its meaning is straightforward. Thus, the transparent Turkish morphemes also render the relationship between the verbs *konuşu-yor-sun* and *konuşu-yor-sun-uz* transparent, which is not the case for the German verbs *spiel-e* (1<sup>st</sup> SG present), *spiel-st* (2nd SG present) and *spiel-t* (3rd SG present). The endings *-e*, *-st*, and *-t* do not share a common marker for singular or present tense; instead, they express TENSE, PERSON and NUMBER cumulatively. The lack of transparency of morphemes is also referred to as opacity (see, e.g., Laaha & Gillis, 2007). Similarly, inflectional classes and irregularity render the relationship between meanings and the forms through which they are expressed non-transparent. It is important to note that exponence transparency, which is tested in this thesis, is just one parameter which influences the transparency of a form or system. However, since only this aspect is tested here, separative exponence and transparency are used interchangeably when referring to the experiments presented in this thesis. Previous studies have not always clearly distinguished between different aspects of transparency, which adds to the difficulty of understanding how transparency impacts language acquisition. More examples of (non-)transparency and examples of studies will be discussed in **Section 2.1**.

#### 1.2.5 Word length

Another striking difference between agglutinating and fusional systems lies in the length of words. This is another consequence of exponence type: if one assumes that both systems express the same number of complex meanings, then it is easy to see

how the agglutinating technique yields longer words than the fusional technique (Plank, 1986; Plungian, 2008). The examples of *spielst* and *konuşuyorsunuz* illustrate these differences, and my artificial language systems reflect these as well. Again, contrasting *konuşuyorsunuz* with *konuşuyorsun*, it also becomes apparent how agglutinating structures re-use morphemes to express different combinations of meaning: in Example (3) from Turkish, all forms in the present tense continuous paradigm of *konuşmak* (speak) contain the morpheme *-(u)yor* to indicate present tense continuous. Example (5) illustrates this compositionality even further: replacing *-(u)yor* with *-ur* yields the forms of the verb in the present simple (van Schaaik, 2020). This mapping of meanings onto segmentable forms is not the case for the German verb *spielen* (play) in present tense as shown in Example (4): since the morpheme for present tense cumulates with those of PERSON and NUMBER, similar meanings do not necessarily coincide with similar forms. This has an effect on the frequency with which forms occur: regular agglutinating morphemes will occur more frequently than typical fusional morphemes, a finding that I will illustrate further in **Section 2.3**.

(3) The verb *konuşmak* in Turkish ('speak', present tense continuous)

1 <sup>st</sup> SG	<i>konuş-(u)yor-um</i>
2 <sup>nd</sup> SG	<i>konuş-(u)yor-sun</i>
3 <sup>rd</sup> SG	<i>konuş-(u)yor</i>
1 <sup>st</sup> PL	<i>konuş-(u)yor-uz</i>
2 <sup>nd</sup> PL	<i>konuş-(u)yor-sun-uz</i>
3 <sup>rd</sup> PL	<i>konuş-(u)yor-lar</i>

(4) The verb *spielen* in German ('play', present tense)

1 <sup>st</sup> SG	<i>spiel-e</i>
2 <sup>nd</sup> SG	<i>spiel-st</i>
3 <sup>rd</sup> SG	<i>spiel-t</i>
1 <sup>st</sup> PL	<i>spiel-en</i>
2 <sup>nd</sup> PL	<i>spiel-t</i>
3 <sup>rd</sup> PL	<i>spiel-en</i>

(5) The verb *konuşmak* in Turkish ('speak', present tense)

1 <sup>st</sup> SG	<i>konuş-ur-um</i>
2 <sup>nd</sup> SG	<i>konuş-ur-sun</i>
3 <sup>rd</sup> SG	<i>konuş-ur</i>
1 <sup>st</sup> PL	<i>konuş-ur-uz</i>
2 <sup>nd</sup> PL	<i>konuş-ur-sun-uz</i>
3 <sup>rd</sup> PL	<i>konuş-ur-lar</i>

Additional typical characteristics of agglutinating and fusional systems have been named in earlier research, but they extend beyond the scope of this thesis and do not directly relate to the question of transparency. In sum, while many characteristics of morphological systems have been pointed out in the literature, the type of exponence is arguably central to the distinction between agglutination and fusion (Bybee, 1985; Haspelmath, 2009; Plungian, 2008). Greenberg (1960) describes that in the traditional sense, morphological types were first and foremost characterised by exponence type, whereby agglutination was the technique in which morphemes and roots were concatenated mechanically, without any elements being modified, while fusion of the forms involved was the counterpart to it. Greenberg adds that “the classic meaning of *agglutination*” is that “meaningful elements are joined with slight or no modification” (p.185). Plank (1986) even considers that there may be language types, as long as they are defined by exponence, not by agglutination or fusion. However, as will be shown below, data from natural languages suggests that this is still too generalising. Indeed, Dressler (2003) observes that even noun and verb morphology of the same language can resemble or develop towards different morphological types. Plank (1999) observes splits between separative and cumulative exponence in various areas of morphology, including within single grammatical features, indicating that natural languages typically display more than just one morphological technique. This point will be taken up in **Section 1.3** below.

Yet, it appears that exponence impacts a number of further features, such as affix variation, paradigm size, word length and the frequency of occurrence of morphemes. Fusional systems have been found to display more stem and affix variation, which naturally leads to more distinct forms, assuming that grammatical meanings are expressed overtly and that syncretisms remain sparse. Fusion and the resulting lack of compositional structure also impact the frequency with which individual morphemes occur. These characteristics concerning the shape and number

of different forms in an inflectional paradigm have been claimed to stand in relation to the size of paradigms, which have been argued can be larger for agglutinating systems (Plank, 1986). In a more general way, Laaha & Gillis (2007) note that agglutinating systems tend to be richer in inflectional morphology, and this appears to be a consequence of compositional structure. Hereby, the first problem with trying to classify whole languages as agglutinating or fusional has been described in detail, by showing that both types comprise a number of different characteristics. However, it is not uncommon that a given language only displays some of the features typical of a morphological technique.

### **1.3 Difficulties in ascribing entire languages to a morphological type**

The second problematic point with the notion of morphological types raised above is whether the theoretical distinction between agglutinating and fusional types in fact fully maps onto natural languages. Firstly, Haspelmath (2009) offers an explanation as to where the idea of whole agglutinating or fusional languages may come from: “In the early days of modern typology, what struck linguists was the differences between languages, and not so much the more abstract differences between patterns (p.16).” A further reason for the survival of the agglutination-fusion distinction could be the fact that comparative descriptions have often rested on very typical exemplars of these types, such as Turkish and Latin (Haspelmath, 2009). However, one needs to ask whether the notion of morphological language types still holds with the typological knowledge of today, given regular criticism (Dressler, 2003; Greenberg, 1960; Laaha & Gillis, 2007). Indeed, considering many morphological descriptions of natural languages, it is clear that languages on the whole often do not consistently display morphology from only one type (Comrie, 1981; Plank, 1999). This finding is also supported by Bybee (1985) who states that mixed cases are more frequent than extreme, typical cases, but finds that strongly agglutinating languages are much rarer than strongly fusional languages.

In fact, many examples of a mix of morphological techniques can be found in natural languages, whereby the ratio of features typical of the different types can differ. Some languages seem quite balanced between two techniques, such as Mongolian and Semitic (in this case balanced between agglutination and fusion) (Plungian, 2008). Other cases lean more towards one morphological type, with certain deviations. Even

Turkish, which is often regarded as a prime example of the agglutinating type, expresses most PERSON and NUMBER forms in the present tense cumulatively, although present tense is expressed with a separative exponent. This means that separative and cumulative exponents are combined in Turkish verbs (see examples (3) and (5) above and example (8) in **Section 2.3.1**). Finno-Ugric languages like Mansi and Udmurt show affix variation, although they are considered agglutinating languages, which typically display affix uniformity (Plungian, 2008). Similarly, Hungarian is considered agglutinating overall, but it uses suffixes indicating possessive PERSON and NUMBER cumulatively, as shown in example (6) below:

(6) Illustration of cumulative expression of possessive PERSON and NUMBER in Hungarian (Uralic; Europe)

1 SG *kez-em*  
2 SG *kez-ed*  
3 SG *kez-e*  
1 PL *kez-ünk*  
2 PL *kez-etek*  
3 PL *kez-ük*  
'my hand etc.'

*(adapted from Haspelmath, 2009, p.20)*

Finnish is also largely considered agglutinating (see for instance Laaha & Gillis, 2007), but Plungian (2008) points out that while it has little cumulation in its nominal inflection, it has lots of grammatically conditioned variation and fusion on morpheme boundaries. Another example is Basque, which mostly expresses NUMBER and CASE by separative exponence, but uses cumulative exponence to express plural absolutive. Estonian, which also expresses NUMBER and CASE largely by separative exponence, uses cumulative exponence for the nominative, genitive and partitive (Plank, 1986). Laaha & Gillis (2007) observe that also those languages in their sample which are classified as strongly fusional display a few isolating characteristics. They refer to an example from Russian, which expresses nominative and genitive CASE through inflections, but expresses locative CASE through prepositions. Finally, example (7) shows how agglutinating and fusional techniques are used to varying extents within the paradigm of a single verb in Greek: Here, the active past punctual forms of the verb 'write' in Modern Greek are much more agglutinating than the highly fusional passive present forms of the same verb (Plungian, 2008).

(7) Verbal paradigm in Modern Greek displaying agglutinating and fusional techniques to varying extents

Modern Greek, the verb 'write':

	3SG	3PL	
Active			
PRES:(DUR)	<i>γράφ-ι</i>	<i>γράφ-ουν</i>	
PAST:DUR	<i>έ-γράφ-ε</i>	<i>έ-γράφ-αν</i>	
PAST:PUNCT	<i>έ-γράφ-σ-ε</i>	<i>έ-γράφ-σ-αν</i>	
Passive			
PRES:(DUR)	<i>γράφ-ετε</i>	<i>γράφ-ονται</i>	
PAST:DUR	<i>γράφ-όταν</i>	<i>γράφ-ονταν</i>	
PAST:PUNCT	<i>γράφ-τικ-ε</i>	<i>γράφ-τικ-αν</i>	<i>(adopted from Plungian, 2008)</i>

The examples above illustrate how languages can display instances of split morphology, where different morphological techniques are found in various areas of the language, for instance between word classes, but also within a single grammatical feature or individual lexeme (Plank, 1999). Mixed paradigms like these make it difficult to assign a language to a single type. Given a wealth of examples of this kind, one can conclude that many languages, if not the majority of them, do not consistently display properties of a particular morphological type. Indeed, this fact has been stated many times. Greenberg (1960) notes that “[a] term like *agglutinative* applies primarily to a single construction”, adding that languages may overall tend towards a certain type, while displaying features of other types as well (p.182). Already Sapir (1921) noted that morphological types are not mutually exclusive between languages. Interestingly, Greenberg (1960) describes how Sapir used labels such as “agglutinative-fusional” to indicate that a language employed both techniques to similar extents (p.184). Similarly, Plungian (2008) observes that languages can display different combinations of characteristics and may only show a subset of those that have been described as prototypical of a certain morphological type.

In an attempt to test the consistency of the agglutination parameter within a language, Haspelmath (2009) finds in a sample of 30 languages that languages are not coherently cumulative or separative in their formatives across nouns and verbs, which speaks against the notion of agglutinating or fusional languages on the whole, if one regards exponence as a major indicator. Thus, this is in contradiction to Plank’s (1986) earlier suggestion that whole languages may be classified based on

exponence. Additionally, Haspelmath (2009) finds that it is not the case that the languages in his sample that display a certain feature typical of one morphological type also display all the other features typical of this type. These findings again suggest that languages do not necessarily exhibit all characteristics, and not necessarily consistently, of a particular morphological type.

Haspelmath (2009) consequently concludes that classifying whole languages as cumulating or separative is not directly helpful, and instead suggests restricting these terms to morphological subsystems. Dressler (2007) proposes that “noun and verb inflection may have a different typological character within one and the same language and develop diachronically in typologically different directions”, adding that French nouns have developed much more from fusion to isolation than verbs (p.4). In a similar vein, Plungian (2008) speculates that intermediate morphological types may be more frequent than prototypical types. This offers an answer to the second area of concern raised above, namely that languages do not neatly adhere to only one morphological type, although such simplifications are sometimes used (e.g., Hengeveld & Leufkens, 2018). However, despite the limitations of such typological classification, it is still highly useful when considering the effects that co-occurring characteristics can have, for instance during language acquisition.

In keeping with the recognition of language-internal variation, this thesis focuses on testing the role that the transparency of the mapping between grammatical meaning and morphological form plays for the learnability of a language, or of a morphological subsystem for that matter. Transparency is intricately linked to exponence, making it a natural focus of investigation. I therefore use the type of exponence as the main criterion differentiating agglutinating and fusional paradigms in the experiments presented in this thesis and use separative exponence mostly synonymously with agglutination and cumulative exponence synonymously with fusion, having noted that the picture in natural languages is much more complex. As described above, natural languages may not use separative exponence consistently across their system and may not display other features typical of the agglutinating type, the same being true of the fusional type. The terms ‘agglutinating’ and ‘fusional’ might therefore not be very meaningful for empirical tests like the one in Haspelmath (2009). However, the artificial morphological paradigm I employ is much more restricted and does not contain most of the intricacies found in many naturally

occurring paradigms. Therefore, I continue to use these terms below to describe the artificial languages used in my studies.

### **1.3.1 Continuous development between morphological types**

A final point to be made about morphological types is that they are in flux (Greenberg, 1960; Sapir, 1921; Slobin, 1980). Albeit a slow development as with most changes, it has often been argued that the morphological make-up of languages changes over time and largely follows a fixed order. On one level, the replacement of synthetic constructions by analytic ones as found in many European and creole languages, such as French *elle chante-r-a* ('she will sing') by *elle va chanter* ('she is going to sing'), has received lots of attention since the 1990s and has been investigated as instances of grammaticalisation or analyticisation (Haspelmath & Michaelis, 2017, p.5). On another level, isolating or analytic systems have been found to develop into agglutinating systems, which develop into fusional systems. Which morphemes fuse together may to a large extent be governed by the principle of relevance, according to which elements which are highly relevant to each other content-wise will also appear near each other (Bybee, 1985). At the end of this process, fusional structures change into isolating structures again, which completes what has become known as the morphological cycle (Croft, 2003; Igartua, 2015; Plungian, 2008). The development from fusion to isolation is well exemplified in the case of English, which is known for its loss of inflections since the Old English period (Kortmann, 2005). Similarly, French has lost inflections, especially in spoken language, and has been classified as a weakly fusional language in its present state (Laaha & Gillis, 2007). Slobin (1980) notes that over the time of 4,000 years, ancient Egyptian repeated two cycles between analyticity and syntheticity. The observation of languages changing between morphological types has been made early on, for instance by Sapir (1921) and Friedrich Schlegel (see Blumenthal-Dramé & Kortmann, 2013).

The change from isolation towards fusion is what Plungian (2008) describes as a "gradual loss of syntactic (and morphological) autonomy of linguistic units" (p.677). He also notes that agglutinating morphemes behave more similarly to words rather than to bound morphemes in the way they combine with each other. Further, the combinability of agglutinating affixes with different word classes resembles that of auxiliaries in analytic languages (Plungian, 2008). These theoretical observations

appear to support the claim of the morphological cycle, where an agglutinating stage also lies between an isolating and a fusional stage on a diachronic level. Indeed, Plungian (2008) finds that

“the type of grammatical systems found in agglutinative languages is closer to what occurs in analytic and isolating languages; at the same time, the morphological structure of the word (namely, the abundant use of affixes) is more like that of flecive languages. Agglutination is intermediate” (p.676).

This brings the theoretical and the historical perspective together, in that agglutinating systems are not only intermediate between two other types in terms of their grammar and word forms, but also an intermediate stage in the process of language change.

Crucially, Plungian (2008) points to a relationship between cumulation and fusion which may help explain why agglutinating structures tend to develop into fusional structures and not vice versa. He states that cumulation and non-uniformity often arise from fusion, whereby neighbouring morphemes assimilate phonologically. In this sense, cumulation may well be the result of phonological changes, at least to some extent (see also Igartua, 2015). Similarly, Bybee (1985) notes that reductive phonological changes leading to allomorphy are common across languages. Adopting a diachronic perspective, Bybee (1985) argues that agglutinating languages may not display this feature since they find themselves in a state prior to the one where fusion occurs. She further adds that semantic notions which are related or relevant to one another are more likely to form a close-knit semantic structure, considering they also need to be learnable. This observation is very helpful to note because it suggests that cumulation is not the driver or target of change, but the product of naturally occurring sound changes if necessary semantic conditions are met. This is compatible with the oft-stated claim that agglutinating structures are easier to learn: if cumulation results – at least to some degree – from phonological changes, then the development of fusional systems is not in conflict with a *learner’s* preference for agglutinating systems. Further, fusional systems arguably have their own advantages, such as efficiency, although this is likely not driven by learners, but by speakers (Kusters, 2003).

In **Section 1.3**, I have already illustrated that at any point in time, a single language most likely displays more than one morphological technique. In addition to this, the present section has highlighted that languages do not only display different morphological techniques at any one point in time, but that diachronic changes in the

morphological techniques they employ have been observed as well, and may indeed be common. This adds to the idea that trying to force a holistic definition to an entire language is less meaningful. Noting that morphological patterns naturally change over time adds to the understanding that patterns and changes within languages may be more revealing than attempting to classify languages in full. In a similar vein, observing changes may provide a window to potential causes of such changes as well.

Another point to raise in this regard is that such implicational statements have been attempted. For instance, Igartua (2015) finds that some developments in the inflectional systems of different languages provide counter-examples to the direction of the morphological cycle. In multiple languages, he identifies instances where agglutinating patterns develop inside otherwise fusional verbal and nominal paradigms. Igartua (2015) relates these developments from fusion to agglutination to the fact that they happen in language contact situations involving second language learning. These findings will become important in **Section 2.1.5**, where I will detail how these exceptions have been interpreted in light of learnability constraints.

### **1.3.2 Geographical distribution**

The difficulties in classifying entire languages as agglutinating or fusional, as described in the previous section, make it complicated or even meaningless to try to assign natural languages to one morphological type. The observation of morphological developments following a particular cycle suggests that the morphology of languages is changing over time and a classification we can obtain today is only a momentous impression, although language change is slow. This is arguably the case with most aspects of languages, but it is important to see this in relation to claims about learnability. Namely, if we were to simply observe that one type is more frequent than another, this does not automatically mean that it is easier to learn, because morphology appears to develop along certain axes and because languages face other pressures beyond the pressure to be learnable (Kusters, 2003; Plank, 1983; Slobin, 1980). Thus, a claim linking learnability to the transparency of agglutinating systems needs to make a more precise case of the causative relation. However, patterns of geographical distributions of agglutinating and fusional subsystems of languages have been observed. I will describe them in the following to show that both morphological

types are well attested among the languages of the world, while there may also be geographical clusters to some extent.

The focus of the early phases of morphological typology may be described as eurocentric, which seems to explain why fusional systems were being associated with highly developed structures superior to those of agglutinating systems in the 19<sup>th</sup> century. However, the empirical finding that many Indo-European languages commonly use cumulative exponence, or even are described as fusional more generally, is well supported today. As Haspelmath (2009) puts it, “[c]umulation is simply a rare phenomenon outside of the Indo-European family” (p.22). However, cumulation has been found to be very common in certain areas of the inflectional system, especially cumulation of PERSON and NUMBER as well as of TENSE and ASPECT, which are less rare in non-Indo-European languages (Bickel & Nichols, 2013b; Haspelmath, 2009; Plungian, 2008). With regards to declension, cumulation of NUMBER and CASE is frequently attested, and can also be found in Indo-European languages (Bickel & Nichols, 2013b; Plungian, 2008). However, it is not restricted to this language family, since Bickel & Nichols (2013b) have found it also in Chukchi, Finnish, Nenets, West Greenlandic, and Yaqui. In terms of overall numbers, monoexponential, that is separative, morphemes have been claimed to be “[t]he universal default” (Bickel & Nichols, 2013b). In comparison to these, polyexponential, that is cumulative, morphemes are understood to be rarer (Bickel & Nichols, 2013b). However, both types of exponence are attested across the world and in different language families.

It is also possible to find languages that have been described as rather typical of certain morphological types in the literature. A prominent example of the agglutinating type is Turkish, but also Finnish, the Finno-Ugric branch more generally, Basque, Swahili and Japanese have been named as representatives of this type (Bickel & Nichols, 2013a; Haspelmath, 2009; Igartua, 2015; Kortmann, 2005; Laaha & Gillis, 2007; Plungian, 2008)<sup>3</sup>. While this suggests that agglutinating-type languages are found across the world, there appear to be certain geographical areas where agglutination occurs frequently. Plungian (2008) describes “most Austronesian

---

<sup>3</sup> Haspelmath (2009) lists further languages which have been described as agglutinating in the literature: Nivkh, Swahili, Quechua, Lezgian, Shoshone, Martuthunira, Evenki, Yukaghir, Hungarian, Kannada (p.24). Plungian (2008) also names Quechua as an agglutinating example.

languages”, “many Australian languages”, Dravidian as well as “most Finno-Ugric” languages as agglutinating (p.670).

Prototypical examples of fusional languages have also been suggested. These include “classical Indo-European languages (Sanskrit, Ancient Greek, Latin), Slavic languages, many American Indian languages” (Plungian, 2008, p.670; see also Haspelmath, 2009, regarding Latin). Russian, Croatian and Greek have also been described as examples of fusional languages (Laaha & Gillis, 2007). German and Dutch are also sometimes presented as fusional examples, similar to French, although these are seen as less fusional and partly isolating compared to more prototypical fusional languages (Kortmann, 2005; Laaha & Gillis, 2007). Based on these examples, it is clear to see that both agglutinating and fusional types exist across different language families and geographical areas, while there also seem to be certain clusters where related languages use similar morphological techniques. Overall, agglutination and fusion are both well-represented across the world, although separative morphemes are understood to be more common.

## **Summary**

This section reviewed the development of morphological typology as a research field and illustrated a classification system based on the morphological make-up of languages. It described core differences between agglutinating and fusional types and pointed out that it may often be more meaningful to distinguish agglutinating and fusional subsystems. Importantly, it has often been claimed that based on their more transparent form-to-meaning mappings, agglutinating structures are easier to learn than fusional structures. Investigating this proposed connection lies at the heart of the experiments presented in this thesis. What is important to consider, however, is that transparency – or at least what has been called transparency in the literature – can result from several factors, including separative exponence, lack of allomorphy, lack of inflectional classes and lack of syncretism. In this thesis, I investigate the role of separative exponence as a measure of transparency which relates to individual forms. Features like lack of allomorphy, etc., what I will call regularity, are features of paradigms (or systems). As I discuss below, these features correlate with transparency in terms of separative exponence, making the latter difficult to study in

isolation. The next section will summarise claims made in the literature about the effect of transparency on learnability of morphological systems.

## Chapter 2

### Learnability differences across morphological types and the impact of L1 morphology

It is generally agreed upon that inflectional morphology presents one of the biggest challenges to L2 learners (Clahsen et al., 2010; DeKeyser, 2005; Prévost & White, 2000). DeKeyser (2005) notes that “even the most basic morphology is often lacking from the speech of untutored immigrants” (p.7). However, all morphology is not created equal, and within the realm of inflectional morphology, it has been suggested that some structures are easier to acquire than others. This also generally applies to L1 acquisition, where some structures are learned earlier than other structures. Forms which are the only carriers of a particular piece of grammatical information have been suggested to be learned earlier than redundant forms such as agreement markers (DeKeyser, 2005). Also factors like semantic complexity, phonological salience, frequency and morphophonological regularity have been proposed as determinants of acquisition order (Brown, 1973; Goldschneider & DeKeyser, 2002; Penke, 2012).

Claims about learnability have also been derived based on morphological types. Certainly, isolating languages by definition are not very rich in inflections and a learner of these languages circumvents the issue of acquiring a large number of inflectional morphology (Comrie, 1981). Interestingly, children acquiring languages with rather sparse inflectional morphology have been found to take longer to acquire their morphological system than when learning a morphologically richer language (Aksu-Koç & Slobin, 1985; De Houwer & Gillis, 1998; Dressler, 2012; Hengeveld & Leufkens, 2018; Laaha & Gillis, 2007; Penke, 2012). However, those languages rich in inflections, that is synthetic languages, have been suggested to differ in learnability based on the morphological techniques they employ. In particular, one major claim is that due to the higher transparency of the form-meaning-relationship, agglutinating morphology is easier to learn, both in L1 and L2 acquisition (e.g., Aksu-Koç & Slobin, 1985; Bittner et al., 2003; Hengeveld & Leufkens, 2018; Igartua, 2015; Penke, 2012).

However, transparency as it relates to learnability is often confounded with several other factors that have also been claimed to influence learning, namely regularity, compositionality, generalisability and frequency. Transparency and regularity will be discussed in detail in **Section 2.1**. Compositionality applies across

different levels of the language and describes the idea that the meaning of a composite form can be inferred from the meanings of its sub-parts (Krifka, 2001). This may apply to the sentence level, but it also applies word-internally, as in the example of the Turkish verb *konuşu-yor-sun-uz* ('you are speaking'). **Section 2.2** illustrates the link between compositionality and morphological types and how compositional systems allow a learner to generalise linguistic forms to new contexts. **Section 2.3** shows how the frequency of morphemes can differ depending on the morphological system in question and how this may impact on learning ease. Beyond intricacies of the system in question, also the relationship between L1 and L2 morphology and potential transfer from the L1 onto the L2 have been investigated as a predictor of acquisition difficulty (DeKeyser, 2005; Gass & Selinker, 2008, Luk & Shirai, 2009). The question of how morphological structures present in the L1 may impact learning of L2 morphology will be elaborated on in **Section 2.5**.

## **2.1 The role of transparency for learnability**

The very nature of agglutinating morphology means that it is more transparent than fusional morphology. This rests to some degree on the fact that agglutinating morphemes typically express grammatical meaning in a separative manner – every unit of meaning is mapped onto a separate morpheme. This makes the relationship between meaning and form very clear. Additional characteristics of agglutinating systems, such as lack of affix variation and lack of inflectional classes, further increase the transparency and regularity of the system. Non-transparency exists in two ways: multiple meanings can be expressed through a single form (e.g., cumulative exponence and syncretism), which results in a one-to-many relationship, and multiple forms can express the same meaning (e.g., multiple exponence and allomorphy), which leads to a many-to-one relationship between form and meaning (Hengeveld and Leufkens, 2018; Kusters, 2003). Since these are typical characteristics of fusional systems, it shows how typical agglutination and fusion are diametrically opposed with regards to the transparency feature. Further, transparency is commonly seen as a highly valuable characteristic of languages in linguistic description (e.g., Bybee, 1985; Don, 2017). In turn, it has been suggested many times that higher levels of transparency increase the learnability of morphological systems both in L1 and in L2 contexts (DeKeyser, 2005; Dressler, 2012; Igartua, 2015; Penke, 2012).

In the next section, I will define the notion of transparency I use here, which will be followed by a detailed description of how agglutinating systems are more transparent and regular than fusional systems (**Section 2.1.2**). I will then move on to theoretical claims that transparent forms are easier to learn in L1 and L2 learning contexts (**Section 2.1.3**). A discussion of studies on natural language acquisition implying transparency as a determinant of acquisition ease will be provided in **Section 2.1.4**. The final part of this section will describe research on language change which also implies transparency as a driving force (**Section 2.1.5**).

### **2.1.1 Transparency as the principle of ‘one meaning one form’**

Morphological transparency can be defined as a one-to-one relationship between a form unit and a meaning unit (Hengeveld & Leufkens, 2018; Kusters, 2003). In the study of inflectional morphology, we are interested in morphs as the unit of form, and morphemes as the unit of meaning. If a morpheme is expressed by one particular morph throughout, and this morph only expresses the one morpheme, then their relationship is maximally transparent. A maximally transparent example from English is the superlative morpheme *-st*, because it only occurs in this form and the form only expresses superlative meaning. This is an instance of a biunique relationship, which is most typical of agglutinating systems (Dressler, 2012; Laaha & Gillis, 2007).

Deviations from transparency occur when the one-to-one mapping does not hold (Hengeveld & Leufkens, 2018; Kusters, 2003). For example, a morpheme can be expressed through multiple forms, as the English plural is encoded through the allomorphs */-z/*, */-s/*, */-ɪz/*, in addition to irregular forms. This is a case of a many-to-one relationship between form and meaning (Hengeveld & Leufkens, 2018). A single morph can also encode different meanings, depending on context. The morph *-t* in the German verb *spiel-t* (play) can mean third person singular and second person plural. Another non-transparent type of mapping is given when one morph expresses multiple morphemes at once, as does the English suffix *-s* on verbs, cumulatively expressing third person, singular, and present tense. Both are examples of a one-to-many relationship between form and meaning (Hengeveld & Leufkens, 2018). When form-meaning mappings are not transparent, they are also referred to as *opaque* (Laaha & Gillis, 2007). Note that while the English verbal suffix *-s* is an instantiation of non-

transparency in one specific form unit, the English plural allomorphs reduce the transparency across the formal marking of the plural feature. Multiple forms need to be learned and it is not always predictable which form needs to be used (compare *fox* – *foxes* to *ox* – *oxen*). Beyond individual forms, this reduces the transparency of the inflectional system more generally, a point that I elaborate on further below.

Instances of non-transparency can be defined at the level of individual form units, but also at the level of the morphological system more broadly. Kusters (2003) identifies four types of non-transparency, namely fusion, fission, homonymy and allomorphy. Note that I regard homonymy within a single inflectional paradigm as syncretism. Fusion and allomorphy have already been described in the examples above, but note how allomorphy can add irregularity to the inflectional system more generally: if the allomorph depends on the phonological environment, or on grammatical aspects, then a learner will need to remember multiple forms and the contexts in which they occur (see Plungian, 2008). This is less regular compared to a single morph that expresses the grammatical feature. In Turkish, the present tense continuous is exclusively expressed through the morpheme *-yor* on verbs, which does not assimilate to its phonological environment, although other areas of the verb in the present tense continuous do so to retain vowel harmony (personal communication with my Turkish informant; see also van Schaaik, 2020). As a result, the grammatical morpheme is fully transparent. The case of syncretism is similar in that it adds opacity and irregularity to the morphological paradigm and even beyond, since one form unit can express different grammatical meanings depending on context (see the example of the German verb *spiel-t* and the ending *-ing* in English, which can express gerund or a nominalization (Dressler, 2012)). However, this thesis is concerned with meaning variation within paradigms, not with examples of the type ‘bank (in the park) – river bank’, which I regard as instances of homonymy.

In comparison, fusion is more concentrated on individual forms in the first place. For instance, fusion of singular and first person does not necessarily affect fusion of singular and third person. However, there are of course many fusional systems that consistently fuse two features, such as PERSON and NUMBER. The Transparency Principle (Kusters, 2003) is also violated when a meaning is expressed through a combination of multiple forms on the syntagmatic level, which Kusters (2003) labels as fission and others as multiple or extended exponence (Igartua, 2015). Kusters (2003) illustrates this in an example from Arabic: In the verb *t-asrab-na* (drink, 2<sup>nd</sup>

person plural female), the information of NUMBER, PERSON and GENDER is expressed jointly by the affixes *t-* and *-na*. Additional factors that render languages non-transparent include ambiguity, redundancy and zero expression, among others (Bybee, 1985; Don, 2017; Hengeveld and Leufkens, 2018). Also in these cases, a single form expresses multiple meanings (ambiguity) or a meaning is expressed jointly through multiple forms (redundancy).

From the examples above, it should be clear that the notion of transparency as it has been used in the literature has two related notions to it: in the first instance, it refers to a one-to-one correspondence of individual form-meaning pairs on the level of individual exponents, but in a second sense, it describes a regularity on the system level in the sense that forms typically have no variants or that variation follows a regular pattern (Comrie, 1981; Greenberg, 1960). How individual forms affect regularity on a more general level can be illustrated in the following example. If we consider the English plural, the plural suffix *-s* in itself is fully transparent, because it is a single form expressing a single meaning. Regarding the plural allomorphs, however, the variation that occurs depending on the stem means that there is variation between */-z/*, */-s/* and */-ɪz/* on the one hand, but also complete irregularity with regards to forms like *ox – oxen* or *sheep – sheep*. Both reduce the regularity of the system of plural inflection in general because a learner will have to remember multiple forms and not all of them are predictable (Greenberg, 1960). Thus, as already indicated above, many aspects of agglutinating systems, such as lack of allomorphy and inflectional classes, result in regularity on the paradigm level.

As the previous examples indicate, languages can and do display non-transparency of various kinds and it has been noted that non-transparent aspects are a frequent, if not the dominant case in natural languages (Anttila, 1989; Don, 2017; Hengeveld & Leufkens, 2018). For instance, Anttila (1989) observes that despite

“many examples of avoidance of homophony in paradigms (analogic resistance to sound change), or therapeutic removal of homophony (by analogy, borrowing, or grammatically conditioned sound change) [, ...] all languages have homophony to different degrees” (p.184).

Similarly, Hengeveld and Leufkens (2018) find that “it is exceptional for them to display a systematic one-to-one relation between meaning and form, i.e. languages are never completely transparent” (p.139). However, what is important to observe is that the non-

transparent features mentioned above are largely those that are described as typical of fusional systems (see **Section 1.2**). It is therefore not surprising that agglutinating systems on the whole have been described as more transparent than fusional systems (e.g., Hengeveld & Leufkens, 2018) (**Section 2.1.2**). In a further step, claims have been made that transparent agglutinating morphology should be easier to learn than fusional morphology (e.g., Aksu-Koç & Slobin, 1985; Penke, 2012). **Section 2.1.3** introduces these theoretical positions while the remainder of **Section 2.1** will discuss studies of natural language acquisition and language change which claim that transparent forms are easier to learn in both L1 and L2 acquisition.

### **2.1.2 Agglutinating structures are more transparent and more regular than fusional structures**

One of the hallmarks of agglutinating systems is separative exponence, which is by design transparent in its form-meaning mapping. However, as shown above, features such as lack of stem and affix variation, lack of syncretism as well as lack of inflectional classes make agglutinating morphology overall more transparent and more regular than fusional morphology. The link between agglutination and both transparency and regularity has been observed many times in the literature. For instance, Greenberg (1960) states that in a typical case of agglutination, “meaningful elements are joined with slight or no modification” (p.185). Igartua (2015) also notes that separative exponents are simpler than cumulative exponents because they follow the transparency principle, according to which form-meaning mappings are maximally transparent. Dressler (2007) observes that ideal agglutinating systems are more transparent than fusional systems and are characterised by uniformity and biuniqueness of form-function mappings, which suggests higher regularity.

A typological study by Hengeveld & Leufkens (2018) further supports the observed relation between agglutination and transparency/regularity. In a typological study, they classify 30 languages as either transparent or non-transparent on the interpersonal, representational, morphosyntactic, and phonological level and they assess the transparency between these levels (i.e. their measure of transparency combines separative exponence and what I am calling regularity). Those languages in their sample that rank highest in their transparency measure are predominantly

isolating in their morphology, while predominantly fusional languages are among the least transparent ones. Agglutinating languages are situated in the middle, which suggests that languages with a large amount of agglutination tend to display more overall transparency/regularity than predominantly fusional languages. While there are exceptions to the tendency in their sample, Hengeveld & Leufkens (2018) find that morphological type correlates with transparency/regularity, even beyond the morphological level.

A consistent mapping between individual forms and their underlying meaning ultimately also leads to regularity across the morphological system. The product of compositional structure, that is the linear combination of morphemes in words, and lack of inflectional classes as well as affix alternation, lack of exception and consistent overt expression results in higher levels of regularity in the way that forms relate to meanings across a morphological paradigm. If non-altering affixes are concatenated into word forms where the meaning of the word follows from the meaning of its parts, then agglutinating words are more predictable than fusional words which are affected by stem and affix variation and by fusion of morpheme boundaries. A learner of a fusional system will have to acquire new cumulative morphemes when a new complex meaning is expressed. Put differently, typical agglutinating systems have fewer exceptions. Further, if affixes tend to be invariant, this reduces the number of different forms and the number of subclasses in inflectional paradigms to be learned.

### **2.1.3 Transparency as a proposed predictor of learnability in theory**

At first, it is noteworthy how the notion of transparency has been viewed in the typological literature. Conceptually, transparent structures have often been regarded as “more ‘natural’, ‘optimal’ or at least simpler than structures lacking such transparent correspondences” (Don, 2017, p.133). Kusters (2003) calls the idea that form-meaning relations should be as transparent as possible the ‘Transparency Principle’, which is one of the principles of inflection, and clearly viewed as an ideal (p.21). In Natural Morphology, language structure is described through a set of principles, which can also conflict with one another and thereby allow language to fulfil neither principle entirely, which is what is typically found in natural languages (Carstairs-McCarthy, 1992). Principles are also seen in need of empirical verification, while guiding this

verification process at the same time. One of the principles proposed by the account of Natural Morphology is transparency, although Dressler et al. (1987) note that “what is considered 'normal' by the speakers of a language depends on far more than on principles independent of the individual language system” (p.61). Nevertheless, transparency is attributed a special status that shapes language along with other factors.

The notion of such transparent relations between meaning and form has been identified early on in linguistic research, going back to Wilhelm von Humboldt and the beginning of European linguistics (Anttila, 1989; Don, 2017). As Bybee (1985) remarks,

“[i]n the literature on morphology, there has always been a great deal of emphasis placed on the optimality of morphological transparency or analyzability or the "one meaning - one form" principle. [...] From the point of view of transparency, regular rules are good and natural, and irregularities, allomorphy and suppletion are undesirable and unnatural” (p.208).

In short, transparency is commonly seen as a highly desirable feature of languages. However, given natural language descriptions, it is not clear to what extent transparency is actually a driving force in shaping morphological systems. In a related notion, Don (2017) demands that transparency needs to “be considered from the perspective of learning strategies rather than from the perspective of the computational system”, by which he means grammar itself (p.133). Nevertheless, transparency is often regarded as an ideal feature of languages and it has often also been named as a predictor of learning ease in both L1 and L2 acquisition (e.g., DeKeyser, 2005; Dressler, 2012; Janssen, 2016; Penke, 2012). Beyond this, the fact that transparent morphemes are a hallmark of agglutinating systems has also led some researchers to claim that agglutinating structures more generally are easier to learn (Aksu-Koç & Slobin, 1985; Igartua 2015; Penke, 2012).

Regarding first language acquisition, Laaha & Gillis (2007) name transparency as one of the core parameters that impact the ease of acquisition of L1 morphology. They also refer to agglutinating morphology such as that of Turkish as being acquired early due to its transparency. Dressler (2012) also lists transparency of inflectional markers, as well as morphological richness in strongly inflecting languages, as factors leading to earlier acquisition in languages in a more general sense. Based on the theory of Natural Morphology, Korecky-Kröll & Dressler (2009) predict that

“morphological patterns which are highly iconic and morphotactically transparent and accommodate biuniqueness are universally preferred and thus should be acquired earlier than less iconic and less transparent ones” (p.268).

Janssen (2016) also highlights another potential benefit of transparent structures, namely that transparent marking can make features more salient in children’s input, which will help them acquire the features earlier. Similarly, Penke (2012) lists the transparency of a form-meaning relationship as one of the predictors of when the inflectional forms in question are starting to be segmented by children, and therefore recognised, in the input, along with typological differences like the one between agglutinating and fusional types. Crucially, she also claims that the factors commonly stated to be influential for the time-course of acquisition, such as frequency, phonological salience, morphological richness, along with transparency and morphological type, interact with each another. This point about a possible interplay of factors will resurface at different stages throughout this thesis.

Although they do not test this in their analysis, Hengeveld & Leufkens (2018) claim that “[t]ransparency, in turn, is an important factor in the learnability of languages”, based on previous literature according to which transparent features in languages are learned earlier and more easily (p.139). Having ranked the 30 languages in their sample according to a transparency hierarchy as described further above, they move on to propose that their level of transparency may impact on rendering whole languages more learnable for L1 learners. While this is a strong claim which resonates with statements found elsewhere, the evidence for the suggested link is missing. It is not clear how transparency in particular areas of the language system translates to learnability, considering that many types of (non-)transparency are discussed, which may be expected to differ in the extent to which they impact acquisition. In addition, transparency interacts with a multitude of other factors, to be discussed in more detail in **Sections 2.2** and **2.3**.

Similar claims about the role of transparency have also been made with respect to second language acquisition. For instance, DeKeyser (2005) proposes that “[r]ather than forms, meanings, or form-meaning relationships, it is the transparency of form-meaning relationships to a learner who is processing language for meaning that determines the difficulty of acquisition” (p.3). He lists opacity as one instance of a non-transparent relationship between form and meaning, for example caused by allomorphy and homophony, both more typical of fusional systems and leading to

irregularity. However, he also attributes semantically complex morphemes like third person singular -s in English to L2 acquisition difficulty and therefore discusses both notions of transparency (DeKeyser, 2005).

Indeed, it should be noted that also regularity is mentioned in the literature as a factor impacting acquisition ease. Kwon (2005) describes lack of exception as one of the oft-suggested determinants of morpheme acquisition order and Don (2017) notes lack of reliability of form-meaning mappings across paradigms as a cause of non-transparency. Dressler (2012) also speaks of the degree to which form-meaning mappings are reliable, along with transparency as a separate factor, which both impact the formal complexity of inflectional forms. In turn, this complexity has an effect on how early and how successfully inflections are learned (Dressler, 2012). Aksu-Koç & Slobin (1985) consider the regularity and consistency of Turkish morphology as one of the reasons for its early acquisition in children. Morphophonological regularity is also one of the predictors of morpheme acquisition order in a study by Goldschneider & DeKeyser (2001). Finally, especially children, but also adults, have been found to reduce unpredictable variation in their input (Culbertson & Newport, 2015; Fehér et al., 2016; Hudson Kam & Newport, 2009). The role of transparency and regularity will be discussed throughout this thesis. In the experiments, the artificial language systems display absolute regularity, but the transparency of individual morphs is being manipulated.

In summary, transparent and regular form-meaning mappings have largely been considered a helpful feature of languages, especially for learners. It is therefore curious to observe that languages are frequently non-transparent in many different ways, as observed by Don (2017) and by Hengeveld & Leufkens (2018):

“[I]t is exceptional for them to display a systematic one-to-one relation between meaning and form, i.e. languages are never completely transparent. Rather, to different degrees they allow ambiguity, discontinuity, and fusion, to mention just a few of the properties that make languages less transparent” (p.139).

Anttila (1989) adds with regards to the transparency principle that

“It has always been known that this principle is a tendency only, like so much in human behavior and biology that is not susceptible to rigorous formulation. No one has ever implied that it actually would lead language to a point where every meaning would have its own form, or total one-to-one correlation between form and meaning. The 'one meaning, one form' principle was also connected

early with psychological factors, which "aim to eliminate purposeless variety" (Wheeler)" (p.107). *(highlight added)*

Brighton et al. (2005) offer a way to reconcile the fact about languages' richness in non-transparent features with the common claim that learners are biased against non-transparent form-meaning mappings. They point out that:

"the end-state of the language acquisition process does not necessarily give us a perfect insight into the biases at play during the process of acquisition. It is possible that child language learners *do* bring some bias against one-to-many mappings to the acquisition of language, but that competing pressures result in an adult competence which contains one-to-many mappings" (p.205).

According to Brighton et al. (2005), the same may hold for many-to-one mappings. These observations illustrate that a potential bias on the part of learners and speakers of a language stands in competition with other pressures. Plank (1983) points to the opposition between transparency of expression and functionality and Hengeveld & Leufkens (2018) state that those features that are least often expressed transparently in their language sample, such as different types of redundancy, may not hinder language processing, or even be facilitative.

In short, many scholars have postulated a link between transparency and learnability. Hengeveld & Leufkens (2018) even allege that L1 learners may find languages more learnable that have overall more transparent features, based on findings that individual transparent and regular features are learned better than non-transparent features. One needs to keep in mind, however, that further pressures shape languages and thus may also influence their learnability (e.g., Kusters, 2003). Trying to reduce languages to their level of transparency may therefore oversimplify a picture that may indeed be highly complex. This may be an aspect that the authors are well aware of but claims that transparency is either the cause for earlier acquisition as shown in **Section 2.1.4** below, or the driver for change, as illustrated in **Section 2.1.5**, have been voiced many times, without necessarily describing the transparency of morphemes in relation to other characteristics of the inflectional system. As I will highlight in **Sections 2.2 and 2.3**, transparent forms are typically also more frequent and more regular than non-transparent ones, which makes them more generalisable (predictable). Without carefully accounting for these co-occurrences, transparency remains a theoretically alleged cause of learning ease, which may deviate from the learner's psychological learning reality (Eddington, 2004). In the studies presented in

**Chapters 3-5**, I investigate the role of transparency in a controlled manner by manipulating the level of fusion in exponents, to see whether more transparent forms are indeed learned more easily, all other things being equal.

#### **2.1.4 The suggested role of transparency in studies of natural language acquisition**

##### **Early studies on morpheme acquisition proposing an effect of semantic complexity and frequency**

Over the past decades, language acquisition research has regularly linked the transparency of morphemes to better learnability. During the 1970s and 1980s, some of the classic morpheme order studies have invoked semantic complexity as one of the determinants of acquisition ease in both L1 and L2 learning. Transparency and semantic complexity rest upon the same notion, they both measure “how many meanings are expressed by a particular form” (Goldschneider & DeKeyser, 2001, p.24, see also Brown, 1976). A morpheme that expresses a single meaning is more transparent and less semantically complex than a morpheme expressing multiple meanings. Consequently, typical fusional morphemes are more semantically complex than typical agglutinating morphemes and a more semantically complex form is predicted to be acquired later than a less complex form (Brown, 1973; Goldschneider & DeKeyser, 2001). In his ground-breaking longitudinal study, Brown (1976) observed the morphological development of three pre-school children acquiring English as their L1. He found that the order in which the 14 grammatical morphemes he analysed were learned was strikingly similar in all three children. Brown (1976) proposed frequency as well as semantic and grammatical complexity as potential determinants of acquisition order. While he was not able to find a correlation between the order of acquisition and the frequency of morphemes in the children’s parental input, he found a correlation between semantic complexity and order of acquisition. For example, *-s* as in *plays* (which expresses 3rd person, singular and present tense) is learned later than *-ing* (progressive). However, Brown (1976) noted that semantic complexity coincided with grammatical complexity in the morphemes studied, as measured by

transformational grammar by Jacobs and Rosenbaum, making it impossible to determine the extent to which each factor affects acquisition order.

Very similar findings were obtained in a cross-sectional study in which de Villiers & de Villiers (1973) investigated the acquisition of the same 14 morphemes as in Brown (1976), but across 21 children between 16 and 40 months learning English. The order of morpheme acquisition in their sample is highly correlated with the one found by Brown (1976). In line with Brown's (1976) findings, de Villiers & de Villiers (1973) do not find any correlation between frequency of morphemes in the parents' speech calculated by Brown (1973) and the order of acquisition among the children in their sample. They do find similar correlations between grammatical and semantic complexity and the order of acquisition, also concluding that it is impossible to determine the extent of each factor's impact. However, de Villiers & de Villiers (1973) suggest to investigate morpheme acquisition order by a combination of the determinants discussed above, together with perceptual salience.

Dulay & Burt (1974) obtained similar findings regarding the order of morphemes in L2 acquisition in children. In a study focusing on the acquisition of 11 English grammatical morphemes, they analysed data from L1 Chinese and L1 Spanish speaking children aged between 6 and 8 years. They find that both groups of children show a nearly identical order of acquiring the morphemes in question and suggest that "universal cognitive mechanisms" underlie the acquisition patterns in the English L2 (p.52). However, unlike the previous two studies, they do not name any particular mechanisms or factors. A number of further morpheme order studies have largely confirmed the acquisition order initially found by Brown (1976) and also suggested possible predictors of this order (Kwon, 2005). Semantic complexity has been named as a factor potentially impacting acquisition order many times (Andersen, 1978; Brown, 1973; Brown, 1976; de Villiers & de Villiers, 1973; Larsen-Freeman, 1976). While some authors reported that they were not able to find such a correlation (e.g., Larsen-Freeman, 1976), others were (Brown, 1976; de Villiers & de Villiers, 1973). Larsen-Freeman (1976) also raises the point that multiple factors may be responsible for the morpheme acquisition order in English that was being so widely observed. In her study, she tests various factors in isolation, among them semantic complexity, but is only able to find a significant correlation between frequency in the input as measured by Brown (1973) and acquisition order.

It may be telling that Larsen-Freeman (1976) lists many plausible factors, such as semantic complexity, phonological complexity and perceptual salience, but is not able to find a significant correlation of a single factor with acquisition order, apart from frequency in the learner's input. This may suggest that frequency plays an important role, and simultaneously, that a combination of factors may determine acquisition ease rather than individual factors. Relatedly, Andersen's (1978) findings from applying an implicational model to acquisition data of 89 Spanish-speaking L2 learners of English (aged between 17 and 19 years) overlap with those of Larsen-Freeman (1976) in that frequency in a learner's input is the strongest determinant of acquisition order. Andersen's (1978) model of second language acquisition also suggests a combination of several factors. He concludes that the interplay between syntactic category, morpheme type (free versus bound) and frequency in the input explains the largest part of the systematicity of the observed acquisition order. However, he speculates that additional factors may play a role, among them similarity to the L1 or L1 transfer, perceptual salience as well as phonological, syntactic and semantic complexity. The likely interplay of several determinants in explaining morpheme acquisition will be discussed using another exemplary study further below.

To summarise, we can note in line with Kwon (2005) that potentially, the main implication of this strand of research in the 1970s was that there might exist a consistent order in which English grammatical morphemes are learned by L1 and L2 speakers with different native languages. Various potential predictors of acquisition ease were discussed as well, with varying success in demonstrating their impact on language acquisition. Generally speaking, correlations were mostly tested by ranking the morphemes based on order of acquisition and correlating this with a rank-order of frequency, semantic complexity and the like, of the morphemes investigated. However, the extent to which researchers elaborated on potential determinants varied considerably and the suggested correlations were not tested in every study (Andersen, 1978; Brown, 1976; de Villiers & de Villiers, 1973; Dulay & Burt, 1974; Larsen-Freeman, 1976). Moreover, correlations were mostly tested between the observed morpheme acquisition order and an order of the morphemes based on one predictor, but not on a combination of multiple predictors. While some studies have found significant effects, others have not. For instance, Brown (1976) and de Villiers & de Villiers (1973) find an effect of semantic complexity in L1 acquisition data, while Larsen-Freeman (1976) finds a non-significant effect in this case for L2 acquisition

data. On the other hand, Larsen-Freeman (1976) finds a significant effect of frequency in the input, contrary to the findings of Brown (1976) and de Villiers & de Villiers (1973). The lack of evidence for a causal effect is why these factors are often referred to as 'putative determinants' (Kwon, 2005).

### **Subsequent studies postulating a similar effect of transparency**

While multiple factors and their combinations were often suggested in the early studies, it is in subsequent research studies that a combination of factors was investigated more rigorously and that transparency as a predictor of acquisition ease has been implicated more directly in discussions of L1 and L2 acquisition data. However, despite frequent suggestions about a positive impact of transparency on learning over the past decades, it remains to some extent questionable until the present day whether transparency is part of the causal mechanism of acquisition. This has to do with the difficulty of examining transparency separately from other factors in inflectional paradigms in natural languages. In this section, I will highlight some studies that exemplify this problem.

In linking to the earlier morpheme order studies, Goldschneider & DeKeyser (2001) conducted a meta-analysis of oral production data of L2 learners' English. They tested whether a combination of factors could explain the order in which over 900 participants acquired English grammatical morphemes. They found that a large amount of the variance in acquisition order can be explained by a combination of the factors of perceptual salience, semantic complexity, morphophonological regularity, syntactic category and frequency. Three of these are related to the morphological structures tested in this thesis. Semantic complexity was operationalised as the number of meanings expressed by a form and therefore corresponds to the notion of transparency used in this thesis. A semantic complexity score was assigned to each functor (similar to the notion of 'morpheme', but also capable of distinguishing instances such as 'past regular' and 'past irregular'), whereby one point was assigned for each grammatical meaning that a functor expresses. For instance, plural -s was assigned a score of 1, third person singular -s a score of 3. The more semantically complex a functor was, the later it was expected to be acquired. Frequency was measured as the number of times a functor appears in the input to the learner, whereby higher frequency is predicted to lead to earlier acquisition. Goldschneider &

DeKeyser (2001) also predict that higher phonological regularity (measured as the number of phonological alternations and the amount of homophony with other grammatical functors) should lead to earlier acquisition. Goldschneider & DeKeyser (2001) emphasise that allomorphy, homophony and fusion all reduce the “salience of the form-meaning relationship”, which should impact learning negatively (p.36). Note also that morphophonological regularity describes transparency on the system level, while fusion relates more to individual forms, which allows them to test both notions of transparency.

Indeed, their multiple regression analysis indicates that around 70 – 80 per cent of the variation in morpheme acquisition order is explained by a combination of the five factors, whereby each contributes a substantial amount, although to differing degrees. These findings suggest that morpheme invariance and lack of homophony, both characteristics more typically found in agglutinating systems, add to acquisition ease, in conjunction with the frequency and transparency of the form, which are also more prominent in agglutinating systems. Overall, their analysis supports claims about the relevance of different factors such as perceptual salience (including syllabicity) and syntactic category (lexical/functional, free/bound) that had been mentioned in earlier research. On the one hand, these findings illustrate that agglutinating morphemes may hold a relatively large variety of benefits, since they tend to be more transparent, regular and frequent than fusional morphemes. On the other hand, some of the factors that could not be shown to have significant effects on learning in earlier studies when tested in isolation, did have a significant effect when tested in combination with other factors. Whether transparency affects acquisition success also individually is the concern of the present thesis.

One further point is important to highlight, namely that while Goldschneider & DeKeyser’s (2001) study constitutes an empirical approach, transparency is still conflated with other factors. Morphemes were scored for each of the five predictors in their analysis, and the model revealed a moderate to high degree of correlation between the predictors themselves. For instance, semantic complexity correlates moderately with frequency, morphophonological regularity and syntactic category, the correlation with phonological salience being somewhat lower. This leads Goldschneider & DeKeyser (2001) to conclude that the effect of a certain predictor in the model may be partly driven by its correlation with other predictors, which means that it is not possible to determine the extent to which each individual predictor

contributes to acquisition ease with certainty. Both notions of transparency identified earlier therefore also correlate with one another in this study: semantic complexity, a measure of item-specific (non-)transparency cannot be assessed separately from morphophonological regularity across the inflectional paradigm. Since there are widespread claims that the form-meaning transparency impacts learnability, it is important to put such a supposed causal relation to the test. This test will be introduced in detail in **Chapter 3**, and it also offers a way of keeping both notions of transparency apart. First, however, a range of further studies suggesting the link under consideration will be presented.

While Goldschneider & DeKeyser's (2001) meta-analysis contributes important insights about potential predictors of morpheme acquisition of L2 English, more recent studies have traced the effect of transparency on L1 and L2 acquisition ease in languages beyond English. Korecky-Kröll & Dressler (2009) conduct a longitudinal study on the acquisition of NUMBER and CASE in a child learning Austrian German, where they observe effects of transparency as well. In general, verbal agreement is mastered earlier than nominal agreement by the child and the authors note that the former is characterised by more transparency and uniformity than the latter. In terms of nominal agreement, the singular possessive genitive marker -s on proper nouns is highly transparent and regular and acquired very early by the child (Korecky-Kröll & Dressler, 2009). Other areas of the nominal inflectional system are less transparent (for example the attribution of GENDER) and less regular (syncretisms between forms are frequent for both NUMBER and CASE). Among plural markers, the last to be acquired by the child is also the least productive, iconic and most opaque marker, which is the umlaut plural (Korecky-Kröll & Dressler, 2009).

Overall, the authors conclude that frequency in the input is the most important predictor for the early phase of acquiring NUMBER and CASE, while productivity, transparency and iconicity are becoming more important in later stages. Thus, they argue for a combined effect of different predictors. However, it is again difficult to trace learnability effects back to individual factors. For instance, in their description of the eight types of plural inflection in German, each of the markers is classified with regards to its productivity, transparency, iconicity and frequency. Often, these are binary classifications such as transparent/opaque, which rest to some degree on the authors' judgments. Yet, alone the fact that every marker exhibits more than one characteristic

makes it very difficult to isolate effects of single factors. This resonates with what has been observed in some of the studies discussed above.

Kempe & Brooks (2008) find that L2 learners of Russian fusional CASE-marking were better able to generalise CASE inflections when they were presented with examples of transparent structures in their input, which served as cues to inflection class membership. The CASE-marking paradigm in Russian is fusional and CASE morphemes simultaneously also encode GENDER, NUMBER and ANIMACY. Syncretism is also widespread in the Russian CASE-marking system. The authors compared acquisition in two conditions, one in which the nominative case of a noun revealed a cue to its GENDER and one in which GENDER was non-transparent in nominative case. In Russian, GENDER is a main cue that determines which declension the CASE inflections belong to. Essentially, the ending of a noun in the nominative predicted genitive and dative endings in the first, but not in the second experiment. In Experiment 1, this regularity should allow for generalisation to novel items. Indeed, Kempe & Brooks (2008) find that generalisation to novel nouns is more successful in Experiment 1 than in Experiment 2, suggesting that reliable cues in the nominative facilitate acquisition of grammatical markers of other cases. The authors interpret this as a positive effect of transparent GENDER marking, but one could also interpret the findings in the sense that transparent marking of GENDER increases the predictability of other forms in the CASE paradigm. This renders the marking more reliable and on the whole more regular, because the phonological and morphological cues of the GENDER marker predict the markers of other cases. Essentially, these findings can be seen as supporting the claim that regularity of a system is helpful for learning and that consistency of marking facilitates acquisition.

Interestingly, Kempe & Brooks (2008) observe that generalisation to novel items is much better in similar studies testing agglutinating systems (Brooks et al., 1993; Frigo & McDonald, 1998; Taraban, 2004, as mentioned in Kempe & Brooks, 2008) and explain this with the fact that the agglutinating nouns consistently contained a transparent GENDER morpheme, such as *bottoo-foo*, *bottoo-ilg*, *glottoik-eef*, *glottoik-rog* where each GENDER morpheme (in this case *-oo* and *-oik*) co-occurred with distinct CASE markers. Such regular and frequent repetitions of GENDER morphemes is not the case in the fusional system tested by Kempe & Brooks (2008), where learners need to infer distributional information, namely that some nouns use one set of inflections and other nouns use another set. Their findings suggest that more explicit

morphological marking is helpful, which lines up with claims about beneficial effects of morphological richness (Dressler, 2012; Laaha & Gillis, 2007). The difference between agglutinating and fusional systems with respect to compositional structure and the frequency of morphemes will be discussed in **Sections 2.2** and **2.3**.

In longitudinal studies of verbal development across languages, Bittner et al. (2003) observe that children detect and reconstruct small verbal paradigms of their language within two to four months after producing their first verbs if their language is morphologically rich and transparent such as Turkish, Finnish, Russian and Croatian. On the other hand, children acquiring languages with less transparent verbal systems, including syncretisms and homophony, are shown to take at least twice as long. Examples of these languages are French on the extreme end and Italian on the less extreme end, with Yucatec Maya, German, Dutch and English in between. These comparisons of the speed of L1 acquisition of morphology across languages suggest that more transparent systems will be learned faster.

Additional evidence for learnability benefits of transparent structures comes from a study on mono- and multimorphemic verbs. Narasimhan & Gullberg (2011) test the acquisition of caused posture verbs in 4- to 5-year-old children speaking either Tamil or Dutch. While Tamil verbs express cause and results of these events separately (equivalent to 'make stand' or 'make lie'), both meanings are conflated in a single morpheme in the Dutch verbs, which makes them less semantically transparent. Crucially, these types of verbs occur much more frequently and in a large variety of contexts in a learner's input in Dutch than in Tamil, where they occur as marked forms in a limited range of contexts. However, Narasimhan & Gullberg's (2011) analysis shows that Tamil-speaking children use the verbs correctly at age 4, matching input frequencies and successfully distinguishing between horizontal and vertical placement events. In contrast, Dutch-speaking children have not fully acquired the verbs by age 5. They do not consistently distinguish correctly between horizontal and vertical placement events and do not match input frequencies to a high degree. These results suggest that semantic transparency drives learning ease more than frequency in the input.

Narasimhan & Gullberg (2011) note that Dutch children may find it difficult to navigate all the contexts in which each posture verb is used. Interestingly, they also observe that Dutch children make very few errors when expressing orientation through intransitive verbs such as *staan* 'stand' and *liggen* 'lie', indicating that they distinguish

the underlying semantics correctly. The difficulty seems to lie in the correct encoding of horizontal and vertical orientation in fusional morphemes. Instead, children produce partly ungrammatical, but semantically compositional, forms by combining the intransitive posture verb *staan* (stand) with causative forms, yielding forms such as *\*doen staan* (make stand) or *laten staan* (let stand). Finally, Narasimhan & Gullberg (2011) observe that children occasionally produce periphrastic constructions encoding orientation and causation and conclude that this indicates that children are aware of the semantics, but have difficulty using monomorphemic verbs which express multiple semantic meanings jointly. Indeed, children's utterances seem to suggest that transparent compositional expressions are easier to process during production. It also appears that transparent form-meaning mappings in Tamil are learned better than frequently occurring cumulative forms in Dutch, which could be linked to the regularity of contexts in which caused posture verbs occur in Tamil. Overall, Narasimhan & Gullberg's (2011) findings lend support to the claim that transparency benefits acquisition because they also consider the frequency of the forms. Their study examines separative and cumulative exponence in lexemes rather than in affixes, however, and it will therefore be insightful to test the role of transparency in affixes under more controlled conditions.

Further claims about concrete examples of the positive impact of transparent structures on L1 and L2 acquisition can be identified in the literature. For instance, early findings by Slobin in the 1970s suggest that transparent structures are acquired earlier in the L1. He observes that "the highly transparent nominal morphology of Turkish is acquired by Turkish children by the age of 2, while Serbian children struggle until the age of 5 with the highly opaque Serbian nominal morphology" (Slobin, 1977, p.190-191, quoted from Hengeveld & Leufkens, 2018, p.170). Hengeveld & Leufkens (2018) also observe discrepancies in the age of acquisition of morphological systems across languages. They refer to the developmental studies by Aksu-Koç & Slobin (1985) on Turkish and De Houwer and Gillis (1998) on Dutch. While Aksu-Koç and Slobin (1985) find that Turkish-learning children master most of their language's verbal morphology by the age of 3 and produce a considerable amount of inflections correctly at the age of 2, De Houwer and Gillis (1998) find that Dutch-learning children at the age of 3 and even beyond have not yet acquired Dutch verbal morphology. This is despite the fact that Dutch is not very rich in inflections, but Hengeveld & Leufkens

(2018) note that it is more opaque than the morphologically rich verbal inflection system found in Turkish.

As noted earlier, claims that the transparency of morphemes impacts acquisition ease have been voiced many times. During the morpheme order studies, transparency was often named, but not consistently tested, and in some cases, the tests did not suggest a significant impact on acquisition. Later research has expanded the focus on languages beyond English. The illustration of examples from studies on natural language acquisition shows how both notions of transparency have been related to acquisition, namely the transparency between individual forms and their meanings as well as a regularity on the system level. However, I have also highlighted some exemplary cases where it is not very clear whether transparency is really the causal factor for learning ease, since it correlates with other factors or because information of further characteristics of the morphological paradigm is missing. This is largely a difficulty of studies with natural languages, where transparency cannot easily be singled out as a predictor. Similar difficulties arise in studies on language change, which have partly suggested transparency as the driver of change. I will highlight some examples in **Section 2.1.5**.

### **Learner preferences for regular structures**

Some of the studies presented above already illustrated that the transparency of individual forms can be difficult to isolate from transparency on the system level (what I consider regularity). In addition, studies have also found that regularity more specifically also leads to acquisition ease. Slobin (1973) notes that Serbo-Croatian-Hungarian bilingual children have been found to have acquired locative case marking in the form of nominal suffixes in Hungarian before the age of 2, while at the same time, their knowledge of locative case marking was only beginning to emerge in Serbo-Croatian, where CASE is marked by a preposition before and a suffix on the noun. While the children clearly had acquired the notion of locative case, their performance differed across both languages. Brown (1976), referring to the same example by Slobin (1973), suggests that the difference in acquisition success may be linked to differing levels of complexity of expressing this feature in both languages. More specifically, Hungarian and Serbo-Croatian differ in the level of transparency with which they mark locative case. CASE-marking in Serbo-Croatian is an instance of

multiple exponence, where CASE is expressed jointly through two elements in the noun phrase, and multiple exponence is an instance of non-transparency as discussed in **Section 2.1.2**. The example therefore provides suggestive evidence for a correlation between transparency and speed of acquisition. Whether non-transparency is also the direct cause of the learning delay in Serbo-Croatian is not investigated by Slobin (1973), however.

Yet, he comments on the second notion of transparency, namely regularity on the system level. Slobin (1973) attributes the earlier acquisition of locative case marking in Hungarian to consistency and regularity of the system, suggesting that overt and regular marking wherever possible is preferred by children. Thus, complexity of a system is not defined by the number of overt markers used. On the contrary, when markers are used consistently, in a regular manner without exceptions or variations, they are considered helpful for learning the morphological system. For instance, children have been found to replace occasional zero marking of morphemes by other morphemes of the same feature across a number of languages, making the marking more consistent (Slobin, 1973). Slobin (1973) also finds that homonymous morphemes tend to be acquired later than unique forms.

Remember that Goldschneider and DeKeyser (2001) also found that morphophonological regularity, defined as the number of phonological alternations and homonymy with other grammatical functors, predicts a large part of the variation in the morpheme acquisition order in L2 English, albeit only in combination with four other predictors. More alternations and more homophony lead to a delay in acquisition. Hengeveld & Leufkens (2018) also highlight that regularisation by children leads to more transparent patterns. Moreover, children have been found to regularise their input across a range of experimental studies and to disprefer unpredictable variation, while adults copy the variability in their input more faithfully (Culbertson & Newport, 2015; Fehér et al., 2016; Hudson Kam & Newport, 2009; Saldana et al., 2021). In short, these studies lend further support to the claim that agglutinating structures should be easier to learn, based on the fact that they exhibit more system-wide regularity.

### 2.1.5 Transparency as a suggested driving force of language change

Transparency has not only been invoked as a factor explaining learning ease, but it has also been postulated as a driver of language change. From the vantage point of the 'ideal' shape of transparency, it is instructive to see how developments towards transparency have been linked to the idea of learnability and cognitive ease more generally. In a chapter on sound changes across languages, Anttila (1989) describes how, in many cases, sound change and analogy work in opposite directions, whereby analogy sets in to 'tidy up' the patterns created through sound change by re-establishing regularities and transparency. For Anttila (1989), transparency as a one-to-one mapping is the most important universal of language and connected to humans' language-learning capacity. For instance, he describes sound changes in Russian, Greek and Estonian where non-transparent correspondences between form and meaning were either prevented or eliminated. Both instances of one form having multiple meanings and instances of meanings expressed through multiple forms were either prevented or replaced. Anttila (1989) claims that the principle of transparency is the underlying engine of such changes across languages and further that the "[a]voidance of homophony and polysemy provide[s] clear evidence of this mental force of 'one meaning, one form'" (p.181). This is one example where a bias against non-transparent structures is implicated as the explanation of language change.

In line with suggestions in the field of language acquisition, Anttila (1989) further links the evidence of a tendency in linguistic changes towards transparent mappings to social and psychological factors, more specifically to instances of language acquisition leading to simplification. In his view, a child learner is guided by the one-meaning-one-form principle and only learns to suppress it as a result of pressure from the speech community. For example, English-learning children will first produce the plural of *foot* as *foots* due to the transparency bias according to which the plural meaning is conveyed by only one form, -s. Based on inacceptance on the part of other English speakers, children will then learn *feet* as an exception to the rule. In other instances, such learner biases may also lead to language change (Anttila, 1989). He further proposes that speakers prefer "a maximally efficient sign system" and that "the mind shuns purposeless variety" (p.181). Anttila (1989) further claims that this bias is at least partly caused by constraints on memory. Overall, he proposes that language change is caused by multiple factors, both language-external (social, psychological,

physiological) and language-internal (phonological). The interplay of different factors, one of them being transparency, is an important notion in this thesis. In fact, Anttila (1989) rejects the idea of searching for a single factor to explain complex developments in language change. However, according to him, the most powerful factor is the psychological preference for simplicity and regularity. The principle of transparency has indeed often been invoked in historical linguistics as an explanation of analogical change (Kusters, 2003).

In a larger language sample, Igartua (2015) describes a rich array of inflectional developments in languages that provide counter-examples to the theory of the morphological cycle. This theory of inflectional change describes a common tendency among languages to develop from an isolating to an agglutinating and subsequently to a fusional stage before becoming more isolating again (Igartua, 2015). However, Igartua (2015) highlights developments from fusion towards agglutination, and thus towards transparency, across a set of languages. Crucially, he links these to cognitive and social factors that may result in a bias for transparency. For instance, he points out that various cumulative markers in the nominal and verbal systems of Armenian, Ossetic, Georgian, Bengali and in Cappadocian Greek dialects, to name a few, have developed into separative markers without an intermediary isolating stage. More specifically, the grammatical categories originally expressed through cumulative exponents in his sample are being fully or partly replaced by separative expression.

Igartua (2015) identifies different mechanisms that might underlie these developments into opposite directions. He describes changes which are in concord with the morphological cycle with respect to agglutination and fusion as language-internally driven, for example through processes like phonetic reduction or erosion. In contrast, he concludes that the trend towards agglutination is for the most part language-externally driven. In his view, it results from processes of morphological simplification which are in line with the transparency principle. This coincides with the claim made by Anttila (1989) with regards to sound changes across languages, namely that transparency is a psychological factor driving language change through simplification. Central to the impact of language-external factors according to Igartua (2015) is the role of language contact involving widespread adult language learning, which he also calls “suboptimal learning” (p.706). As he observes, “shifts from cumulative to separative exponence in the majority of systems studied here seem to have occurred with suspicious frequency in conditions of language contact” (p.707).

Accordingly, simpler transparent structures are supposed to be learned more easily in an L2, causing the observed changes in his sample. Claims about a preference for less complex morphology in adult L2 learners have been made several times, but the extent to which agglutination spreads across neighbouring languages needs further investigation (Lupyan & Dale, 2010; Atkinson et al., 2018).

However, some support for Igartua's (2015) claim, at least with respect to the changes in Ossetic, comes from a study by Plank (1986): in an investigation of the expression of CASE by morphological means in a sample of 47 languages, Plank (1986) observes that languages which predominantly use cumulative markers have an average of 5.6 CASE terms (possible values of the feature CASE), while languages with predominantly agglutinating markers have a higher average, namely 8.6 terms. While no language using cumulative means to express CASE in the sample has more than 10 terms, this threshold is crossed by a number of languages that exhibit separative exponence, which, according to Plank, leads to the higher average of 8.6. Interestingly, Ossetic is part of the sample and falls in the category of 7-10 CASE terms. Plank notes that "it is in principle possible for Case paradigms to expand diachronically, a kind of development exemplified by the acquisition of one or two further Case terms in Ossetic and Lithuanian, in Ossetic concomitantly with the exchange of cumulative for Separatist Case exponents" (p.40). Thus, one might ask whether the shift from cumulative to separative markers in Ossetic is not only due to language-external factors, as Igartua argues, but at least partly due to the fact that it approaches the threshold of 10, which might stand in relation with constraints on memory. It is possible that separative exponence is preferred in general once a paradigm becomes large, and that this preference applies both in first and second language learning, explaining both the data in Plank (1986) and in Igartua (2015).

The observation that agglutinating systems tend to be richer in inflectional morphology than fusional systems has also been stated by Laaha & Gillis (2007), and likely stands in relation to the higher level of structure of agglutinating morphology. This allows agglutinating systems to express complex meanings with fewer unique morphemes than fusional systems. A study by Laaha & Gillis (2007) supports this consideration, where they measure the number of categories, corresponding to Plank's (1986) notion of terms. In their sample of 9 languages, the strongly fusional and strongly agglutinating languages express more form-type categories on nouns and verbs than the weakly fusional languages, but the first two types also differ from

one another: in the strongly fusional group, Russian expresses 12 categories on nouns, Croatian 14, and Greek 7; on verbs, Russian expresses 15 categories, Croatian 18, and Greek 15. In the strongly agglutinating group, Turkish expresses 15 categories on nouns, Finnish 14, and Yucatec Mayan 5. On verbs, they express 37, 19 and 30 categories, respectively. Although the sample only contains 9 languages, there appears to be a certain trend: within-group variability seems similar in the fusional and agglutinating groups for nouns, but when considering verbs, the fusional group expresses between 15 and 18 categories, while the agglutinating group expresses up to twice as many. This aligns with Plank's (1986) finding that the largest number of categories tends to be expressed by agglutinating languages, while smaller numbers of categories can be expressed well by both types of inflection.

However, Igartua (2015) also suggests that agglutinating structures per se are easier to learn, at least in second language acquisition, because they are less complex and more transparent. Indeed, his diachronic observation appears to support the oft-stated assumption that transparent agglutinating structures are easier to acquire. He therefore links the observed morphological changes to a simplification process involving an increase in transparency, which supports second language learning. Notably, however, while language contact is indeed prevalent in the languages and dialects studied in his sample, Igartua (2015) fails to offer a direct causal implication from this learning situation to a preference for transparent structures. While it is plausible that language learners impose certain biases on emerging language structure, it is not directly clear from this study whether it is really a bias for transparency that underlies the observed changes or a bias for other features of agglutinating systems. The issue becomes even more striking given that many other studies are similarly missing this empirical link or fail to prove it statistically. As **Sections 2.2** and **2.3** will highlight further, transparency is also an indicator of other characteristics of markers in inflectional systems, making it even more important to test transparency more rigorously in isolation.

What is additionally interesting to observe is that language change towards more transparent patterns is not always being linked to learner biases. Part of the languages in Igartua's (2015) language sample are Asia Minor Greek dialects, which show a development of agglutinating patterns inside of the nominal NUMBER/CASE paradigm. In another study of the same changes towards agglutination in several Asia Minor Greek dialects by Revithiadou et al. (2015), these changes are merely described

by their underlying language-internal mechanism, namely a reanalysis of the morphological structure of the older fusional paradigms. For instance, in some varieties, the former fusional nominative/singular marker has been reanalysed as a pure NUMBER marker, indicating singular, and a second separative marker for plural has developed. Similarly, CASE is marked relatively distinctively, with zero marking in nominative and accusative and another separative exponent in the genitive (Revithiadou et al., 2015). Second, they propose that the reanalysis was triggered by the levelling of nominal and accusative CASE forms. They observe an influx of Turkish loan words, which they interpret as an additional factor leading to the observed change, since these loanwords exhibit agglutinating structure. Crucially, however, changes towards agglutination are interpreted as a combined effect of mechanisms within Asia Minor Greek dialects and the effect of loanword structure, which are both language-internal. There is no mention of an implication of direct language-external factors, although the authors refer to the additional role of bilingualism and contact-induced language change, which in their view requires further research. They conclude that the language-internal mechanisms described above play the most central role in these changes, contrary to the analysis in Igartua (2015), which might suggest that claims about language learning effects may need careful proof.

Similarly, Sims-Williams (2016) observes that cases of a reduction of lexical allomorphy in Greek can be well-explained by properties of the Greek paradigms and their predictive structure. She concludes that language-internal and language-specific properties “do not rule out the possibility that there are other factors guiding change, but they explain the facts well enough to place the burden of proof on those who claim so” (p.23). This thesis very much supports this perspective: claims about the beneficial role of transparency during language acquisition have been voiced many times, but the burden for research seems to lie in accumulating evidence for the suggested causal relation.

In summary of **Section 2.1**, a transparent mapping between form and meaning is a central characteristic of agglutinating systems. Studies have suggested that transparent features are more easily learned in L1 and L2 acquisition and **Section 2.1.5** illustrated additional attempts in the literature to trace changes in the morphological system of languages to an underlying transparency principle, or a learner preference for transparent structures. In both studies, the authors motivate their analysis with learner biases, but do not establish a concrete connection between

the presumed bias and actual linguistic change. In this sense, the two studies fall in line with many of the studies reviewed in **Section 2.1.4**, where the proposed link is only partly clear. Other studies on similar cases of language change appear more cautious or critical of the suggested impact of transparency. It appears indeed very difficult to observe the impact of transparency in isolation, since it stands in strong connection with further factors. I will describe two factors in detail in the subsequent two sections, namely the compositionality and frequency of inflectional morphemes.

## **2.2 The role of compositionality and the generalisability of agglutinating systems**

### **2.2.1 Definition of compositionality**

One of the characteristics by which agglutinating morphology differs from fusional morphology is its compositionality. This feature is tightly connected to the separative exponence of agglutinating systems and the concatenation of morphemes to form words in order to express more complex meanings. The notion of compositionality goes back to Gottlob Frege (1848-1925), a German mathematician and philosopher, and can generally be understood in the following way: “The meaning of a complex expression is a function of the meanings of its immediate syntactic parts and the way in which they are combined” (Krifka, 2001, p.152; see also Szabó, 2000). A sentence like *Anne gave the book to Tom* follows the definition of compositionality and its interpretation is different from the sentence *Tom gave the book to Anne*. Human language is largely compositional, although clearly not every linguistic utterance is. For instance, think of idioms as in *He layed his cards on the table and disclosed his motives for selling the house* or the oft-cited example *kick the bucket* (Vincent, 2014, p.118). Compositional structure is also attributed to different levels of language. First and foremost, human language is seen as compositional in the sense that sentences are composed of elements like words or phrases (Szabó, 2000; Smith, 2018). However, the elements could also refer to “affixes, words, categories, constructions, idioms, set phrases” (Vincent, 2014, p.3). Accordingly, also the morphological level can be compositional and it often is in the case of synthetic languages (Smith, 2018; Kirby et al., 2015).

The English verb *walked* is compositional because the meaning of the stem *walk* combined with the past ending *-ed* signals the past action of the verb. The past tense verb *taught* is not compositional because it does not consist of sub-units of meaning. Similarly, the sequence of the phonemes of *w-a-l-k* does not constitute a compositional meaning since the individual sounds do not carry independent meaning. An agglutinating word like *duy-du-m* in Turkish (*hear*–past–1<sup>st</sup> person singular, “*I heard*”) is compositional because its overall meaning can be inferred from the meaning of its morphemes, similar to examples (3) and (5) from Turkish in **Section 1.2**, where present tense continuous and present tense simple are expressed as two different morphemes, which combine with verb stems and additional suffixes. Thus, we can understand compositionality in a more general sense, operating at different levels of the language (Vincent, 2014).

Concerning the morphemic level, since transparent morphemes are typically also highly regular (lack of stem and affix variation and of fusion), and combine linearly to form words, agglutinating systems exhibit compositional structure. On the other hand, fusional morphemes are holistic, since the whole form expresses the whole complex meaning. For instance, the verbal suffix *-s* in English cannot be further segmented formally, it expresses PERSON, NUMBER and TENSE jointly. The following section illustrates how compositional structure has been shown to emerge under a pressure for learnability in numerous iterated learning studies, which lends support to the idea that compositionality, and by implication, agglutinating structures, are preferred over fusional structures by learners (e.g., Beckner et al., 2017; Hengeveld & Leufkens, 2018; Igartua, 2015; Kirby et al., 2008; Kirby et al., 2015). It also describes how compositionality allows a learner to generalise the language to unseen signals, an advantage that does not exist in fusional systems (Brighton, 2002). **Section 2.3** introduces a second advantage of compositional systems, namely the higher frequency with which its morphemes occur. Both of these factors need to be accounted for, since they are closely connected to the transparency of agglutinating systems and raise the question to what extent we can assess the impact of transparency separately from these factors.

## 2.2.2 Benefits of compositionality and generalisability during language learning

Having related agglutinating systems to compositionality, it is instructive to see how compositional structure has been shown to arise under learnability pressures in computational and experimental iterated learning studies. Here, learners first acquire a miniature artificial language, which they subsequently need to either recall during testing or use in communication with other participants. The output, including any deviations from the original input language, is then passed on to new generations of learners. In this way, the development of natural languages through a process of repeated learning is imitated (Smith, 2018). Compositional systems have also been shown to be more stable over time, which indicates that they are learned better by subsequent generations. This has been found in computer-based models as well as in iterated learning studies with human participants, both of which will be illustrated in **Section 2.2.3** (Beckner et al., 2017; Chaabouni et al.; 2020, Kirby et al.; 2008, Kirby et al., 2015; Vogt, 2005). Raviv et al. (2019) also find that compositional structure emerges during pair-wise interaction in closed group experiments. What is important to understand from the examples in the following is that compositional structure seems to be preferred in certain learning situations, which provide analogues to the acquisition of natural languages. In other words, these experimental and modelling results lend support to the more general claim that agglutinating systems are easier to learn.

To understand why compositionality increases learnability in these designs, we also need to see how compositional systems allow for generalisability. If we again consider example (3) from Turkish for a moment, we can note that the stem *konuş-* is followed by the suffix *-yor* in each of the verb forms. This suffix expresses present tense continuous in Turkish, and all of the verbs in the example are in the present tense continuous. In other words, each time the meaning unit of present tense continuous is expressed in a verb, we will find the corresponding form unit, *-yor*, in the verbs as well. Assuming that a language is consistently compositional and the mapping between meanings and forms is regular (i.e. there is no allomorphy, etc.), the formal expression of a complex meaning can be predicted based on the forms for the individual meanings, which can be combined into more complex words. As such, if a

language expresses ANIMACY and NUMBER on nouns, and a learner has observed a word expressing animate singular and a word expressing inanimate plural, the learner is able to express inanimate singular since the individual form units have been observed before, and can be recombined. In a fusional system, by comparison, this generalisation is not possible, since animate singular would be expressed as one morpheme and inanimate plural as a different morpheme, and there is not necessarily a predictable formal overlap between these two morphemes and a third one expressing another complex meaning (in this case inanimate singular). In a fusional, or holistic, system, every combination of meanings needs to be observed individually until the system is fully acquired.

The advantage of compositional systems therefore is that a learner does not need to observe the full language (i.e. all possible combinations of grammatical morphemes in our case), it is sufficient to observe all values of all features, such as the singular and plural morphemes of the feature NUMBER and the animate and inanimate morphemes of the feature ANIMACY. This means that in principle, having observed an ideal case, for instance signals for animate\_singular and inanimate\_plural (and the grammatical order of their values in the words) early on in a compositional system, a learner can already generalise the full system (Brighton, 2002). In contrast, in a holistic system, all complex meanings (i.e. all of the fusional morphemes) need to be observed individually until the system is fully acquired. In our example, this means that a learner needs to observe all four fusional morphemes, whereas observing only two compositional words can be sufficient if these contain all four possible values. Every additional feature and every additional value per feature increases this numerical discrepancy between compositional and holistic systems, since new feature values are observed at least as fast as new fusional morphemes (i.e. complex meanings) (Brighton, 2002). This is why a learner can express more meanings after the same amount of exposure to the language if they are acquiring a compositional system. From a learner's perspective, compositionality would therefore be the most optimal structure of the grammar of a language, assuming that a learner has limited memory capacity and wants to use a language that has distinctive signals for distinctive meanings (Kirby et al., 2015). However, the structure of the meaning space is important: Brighton (2002) finds that the greatest advantage of compositional over holistic systems is observed when the number of features is large and the number of values is small. Relatedly, Raviv et al. (2019) refer to the role of input variability in

their experiment: when variability is higher, structured languages are more likely to emerge, which they link to the finding that learners generalise once there are too many variants in the input to memorise.

This link between compositionality and generalisability allows us to see why compositional systems are learned better in the studies described in the following: learners typically receive some input in the form of meaning-signal pairs, such as words and images, during training. In the testing phase, they need to express familiar meanings (i.e. images, etc.), but also unfamiliar meanings which they have not seen during training. In other words, learners need to generalise the language to novel situations (see Beckner et al., 2017; Brighton, 2002; Kirby et al., 2008). Since holistic systems can only express those meanings that have been observed before, compositional systems have an advantage because they can also be used to express previously unseen meanings. Expressivity hereby describes the possibility to express meanings through signals which successfully distinguish objects from one another. Learnability means that the language can be learned from the given input, which is usually a subset of the whole language (Kirby et al., 2015). Pressures for expressivity and learnability can be introduced through the study design, as will be shown below. Iterated learning studies can therefore provide insights into the types of languages that develop as a result of different pressures.

### **2.2.3 The development of compositionality in iterated learning studies in the lab and in computational models**

For instance, Kirby et al. (2015) find that compositional structure arises from the combined pressures of learnability and expressivity in an agent-based iterated learning model as well as in a study with human participants. The pressure for learnability is introduced through a transmission bottleneck in the agent-based model and through the use of transmission chains in the experiments, which means that the language is repeatedly being passed on to naïve learners with no prior knowledge of the language. Expressivity is enforced through the need of participants to communicate information. The size of the bottleneck defines how large the observed subset is in comparison to the whole language (Brighton, 2002). A language that is

learnable is one which allows a learner to express previously unseen meanings based on the exposure to a subpart of the language. As described above, this is theoretically more achievable when the language is compositional, and it is supported by Kirby et al.'s (2015) findings.

In their two studies, smaller bottlenecks indeed lead to more compositional languages and transmission chains of human learners increase language structure, in contrast to a second condition in which the language is learned in a closed group and the same participants repeatedly communicate with each other. In the latter case, the developing language is holistic, which fulfils the expressivity, but not the learnability, function, since the pressure for learnability is lower. If a language is introduced to new generations of learners more often, the pressure for learnability is higher. These findings suggest that compositionality might be a consequence of the need to be learnable, along with being expressive (Kirby et al., 2015). Further, Kirby et al. (2015) argue that “languages which permit the formation of compressed mental representations are easier to learn than those which do not” (p.88), while noting that natural languages are both compressible and expressive.

When testing learning over 10 generations of single learners who are exposed to a subset of the language, Beckner et al. (2017) also observe an increase in compositionality where participants tend to build on the progress of earlier generations in developing language systematicity. Importantly, due to the methodology used in their study, compositionality arises as a product of individual learning, not as a product of direct communication between learners, since the test responses of one participant formed the training input for the next participant. Similarly, in Kirby et al. (2008), compositionality arises purely from learning, since participants were not aware that their linguistic output was being passed on to new generations of learners. This suggests that learning in and of itself can increase compositionality, when the input already displays some degree of structure. However, this increase in compositionality is followed by a plateau effect by the 7<sup>th</sup> generation in Beckner et al. (2017), similar to their result of a re-analysis of the data from Kirby et al. (2008), which even suggests that compositionality decreases over the final generations. Beckner et al. (2017) conclude that despite participants being trained on at least partly compositional languages, this structure can be dropped by participants, suggesting that compositionality is not easy and automatic. This is in line with the finding from Kirby et al. (2015) that compositionality does not arise when there is only a pressure for

learnability or only a pressure for expressivity (see also Smith, 2018, for the same claim). However, independent of whether compositional structure arises from a learnability pressure alone or in combination with a pressure for expressivity, compositionality generally appears to increase learnability.





Similarly, regarding the development of compositionality in iterated learning experiments, Smith (2018) observes that:

“[c]ompositional systems constitute a trade-off between the partially competing pressures from learning and communication, in that compositional grammars are relatively simple yet expressive: the regularities they contain can be exploited by learners during acquisition, but they nonetheless allow meaning to be unambiguously encoded and decoded” (p.156).

This is being compared to a situation in which the same learners communicate repeatedly, the pressure for learnability being lower, and where holistic structure emerges (Smith, 2018). These findings signal that learnability plays a crucial role in the development of compositional structure, namely that new learners find compositional signals easier to acquire.

This speaks to the common claim in the acquisition literature that agglutinating structures are easier to learn during L1 and L2 acquisition (Aksu-Koç & Slobin, 1985; Bittner et al., 2003; Penke, 2012). Natural language learning situations are similar to those in iterated learning experiments in that learners acquire an already existing language, either one developed by other participant pairs or the language of the speech community. It appears that the recombination of elements and the corresponding regularity of compositional systems are factors that assist learners during acquisition.

The paradigm shown in **Figure 2** exemplifies compositional structure as it emerged in an iterated learning chain in Kirby et al. (2015): the sequences ‘mega’ and ‘ege’ encode shape, while ‘wawa’ and ‘wuwu’ encode colour. These forms are recombined in order to express the different possible complex meanings shown in the images. The paradigms used in the artificial language learning experiments presented in this thesis are similar, although the examples shown here stem from a larger paradigm used in Kirby et al. (2015).

	shape1_color1	megawawa
	shape2_color1	egewawa
	shape1_color2	megawuwu
	shape2_color2	egewuwu

**Figure 2:** Compositional language developed in an iterated learning experiment in Kirby et al. (2015) (*adapted from Smith, 2018*)

Comparable results regarding the development of compositionality during language learning have been obtained in agent-based iterated learning models simulating the role of children. Vogt (2005) allows children (learners) in an iterated learning model to talk to each other while still learning the language themselves, which is an instance of horizontal transmission. This creates situations where a child is trying to communicate a meaning whose corresponding signal it has not heard before, which equates to a transmission bottleneck. Vogt (2005) finds that when the majority of speakers in the model are children, compositionality tends to remain more stable than when the majority of speakers are adults, suggesting that children make use of the regularities in their input, since they only observe a subset of the language. This expands the communicative context under which compositionality has been shown to develop as a result of language learning. It may also constitute an important analogue to natural language learning, since children may sometimes want to communicate meanings they have not heard expressed before. In this instance, regularity in the way meaning spaces are encoded aids this purpose (Vogt, 2005). The role of children in shaping language structure is also noted in Slobin (1973), who states that “one could argue that human language could not be so defined if it were not so defined by children, because, in a profound sense, language is created anew by children in each generation” (pp.179-180).

Further evidence for a link between compositionality and better learnability comes from deep multi-agent simulations. Slightly contradictory to Kirby et al. (2015) and Smith (2018), Chaabouni et al. (2020) find that compositionality may even occur by chance instead of through a pressure for learnability, but that “once present, it will survive and thrive, as it guarantees that languages possessing it will generalize and will be easier to learn” (Discussion section: What is compositionality good for?). There is no significant correlation in the other direction – generalisability does not strictly entail compositionality, since there are also other types of structures that allow a learner to generalise but are not compositional. However, the authors observe a strong

positive correlation between the degree of compositionality and the speed of acquisition, as well as with the degree of generalisation accuracy. In other words, the higher the degree of compositionality in a language, the faster it will be acquired. Compositional structures evolving in their simulations also guarantee generalisation to previously unseen composite concepts (Chaabouni et al., 2020). These findings suggest that regardless of the reason for which compositionality emerges, it appears to hold learnability and generalisation benefits. This, according to the authors, makes it “highly desirable”, since it results in better learning by a range of different types of artificial agents. As such, Chaabouni et al.’s (2020) results contribute to the prediction that compositional or agglutinating morphology will be easier to learn than non-compositional, such as fusional, morphology.

That compositional languages can be more stable than holistic languages over time has also been demonstrated in a computational model by Brighton (2002). He investigates the conditions under which compositional structures are more stable and therefore learned better by subsequent generations. In his study, the relative stability of compositional systems is found to be largest when the bottleneck size is small, requiring a large amount of generalisation by the learner. In his model, a learner of a language has four options to produce a signal for a meaning: if the signal for the meaning has been observed before, the learner can produce the same signal. If it has not been observed yet, the learner can either generalise based on the observations that were made, or invent a new signal if generalisation is not possible. In the latter case, the learner can also choose not to produce any signal. Importantly, Brighton (2002) notes that a learner in the model never observes the full language, but always only a subset of it, an analogy to the poverty of the stimulus that was also implemented in the experimental designs discussed above. However, if generalisation is possible and effective, an agent can produce signals even for previously unseen meanings. This raises a crucial point: the possibility to learn the language based on observing only a subset depends on the type of system, i.e. whether it is a compositional system and to what extent.

In Brighton’s (2002) case, these particular conditions of a small bottleneck do lead to compositional systems, which is different from the results of Chaabouni et al. (2020) where a pressure for generalisation does not automatically lead to compositionality. However, compositional systems in Brighton (2002) also have much larger affordances when the system in question is complex, namely when it contains

several features and values. There are limits to complexity as well – when there are too many features and especially too many values, these forms are unlikely to be observed across many object labels, and compositional systems lose their advantage. Overall, Brighton (2002) finds that compositional languages are more stable than holistic languages over time, and therefore learned better, when a relatively small subset of the language is observed and when the language consists of many features with a few values each. This maximizes the generalisation advantage of the compositional system.

Brighton (2002) explains this difference between both systems by the fact that every time an object with its associated signal is observed, feature values are observed as well, and especially when the bottleneck is small, the likelihood of observing all feature values is much higher than the likelihood of observing all objects, which would be required to learn a holistic system. This difference is important – for instance, in Vogt's (2005) iterated learning model, all of the 120 objects are only encountered after around 1,000 trials. This enhances learning of a compositional system and allows for better generalisability. As Brighton (2002) observes, the difference is that “[e]xpressivity is a function of object coverage for holistic language. Expressivity is a function of feature value coverage for compositional language” (p.48). The same conclusion is reached by Vogt (2005) who illustrates that compositional languages can be maintained despite the presence of a bottleneck if the observed utterances contain all possible sub-signals in the language. A holistic language can only be maintained if the unobserved meanings are guessed correctly by the learner, which is highly unlikely.

In short, it has been demonstrated that compositional structure emerges and is better learned in mathematical and agent-based models as well as in iterated learning studies with human participants. It has also been shown that compositional systems are more stable over time, which suggests that learners have a bias for this type of structure, although Brighton (2002) shows that this advantage does not persist under all circumstances. Additionally, it is possible to generalise a compositional system in circumstances of relatively little exposure to the language and compositional languages are more systematic (Szabó, 2000). These features of compositionality suggest a learning advantage of agglutinating over fusional morphology also in natural languages, since typical agglutinating morphology is by default compositional. The fact that results from a range of studies on compositionality support the argument that

iterated learning is a crucial component in the evolution of compositionality aligns with the observation that agglutinating structures are learned better in adults and earlier in L1 acquisition. This may partly be due to the fact that less exposure to the system is required to learn it. Note that this predicts a learning advantage of agglutinating systems more generally. However, compositionality, generalisability and transparency are correlated in agglutinating systems in natural languages and also in some of the artificial languages in experimental studies. A question to be answered therefore is whether learning advantages also emerge when testing these factors in isolation.

This section has shown that agglutinating systems exhibit compositional structure due to the concatenation of invariant morphemes to express complex meanings. In turn, compositionality allows for generalisation on the part of the learner, which means that previously unseen objects can be described after observing only a subpart of the language. This is not possible in typical fusional systems, where complex meanings are not expressed through formal sub-units, but through fusional - or holistic - morphemes, which all need to be observed. A range of studies using computational and participant-based iterated learning techniques have shown that compositionality arises under a pressure for learnability or a pressure for learnability paired with a pressure for expressivity, as is typical of communicative situations. In other words, compositionality aids language acquisition, which may provide another reason why agglutinating systems might be more learnable. However, it also highlights that the transparency of agglutinating systems correlates with compositionality, which entails benefits of its own. One of them is generalisability, as shown above. Another benefit is the higher frequency with which morphemes occur compared to fusional systems, which is the topic of the subsequent section. Studies on the effect of transparency therefore need to account for this interplay, or separate transparency from other factors, which is the approach of this thesis.

## 2.3 Effects of the higher frequency of agglutinating morphemes

### 2.3.1 Agglutinating morphemes occur with a higher frequency than fusional morphemes

Apart from being transparent and compositional, inflectional paradigms displaying agglutinating structures also have numerical advantages. These stem from a difference in the way that meaning is mapped onto form and leads to differences in the number of signals (in our case morphemes) needed to express all possible complex meanings in the system. Since compositionality entails that morphemes are ‘recycled’ to construct a range of different, more complex meanings, they necessarily repeat. This means two advantages of agglutinating systems: they typically contain fewer distinct morphemes, which in turn repeat more often. In example (8) from Turkish below, the morpheme *-r* (present tense) repeats across verb forms. Also the morpheme *-sin* is repeated to express second person. The meanings of these verbs follow from the composition of their morphemes, namely the stem *oyna-* plus present tense plus the respective PERSON and NUMBER morpheme. The latter two features are partly expressed cumulatively and partly separatively (namely in the 2<sup>nd</sup> person singular and 2<sup>nd</sup> person plural).

In example (9) from Spanish, this is not the case: there is no separate morpheme encoding present tense and the six verbal suffixes are all different – neither do morphemes independently encode PERSON or NUMBER. These examples illustrate how individual morphemes in a typical agglutinating system occur more frequently in a learner’s input than individual morphemes in a fusional system (consider the present tense suffix *-r* in (8)). This discrepancy in frequency of occurrence between agglutinating and fusional morphemes applies if we assume the same meaning space in both systems, overt expression throughout and if morphological paradigms do not use syncretisms. The discrepancy also increases when a system has more inflectional classes or allomorphy, since they increase the number of distinct morphemes.

(8) Present simple forms of the Turkish verb *oynamak* ('play')

1 <sup>st</sup> SG	<i>oyna-r-im</i>
2 <sup>nd</sup> SG	<i>oyna-r-sin</i>
3 <sup>rd</sup> SG	<i>oyna-r</i>
1 <sup>st</sup> PL	<i>oyna-r-iz</i>
2 <sup>nd</sup> PL	<i>oyna-r-sin-iz</i>
3 <sup>rd</sup> PL	<i>oyna-r-lar</i>

(9) Present tense forms of the Spanish verb *buscar* ('search')

1 <sup>st</sup> SG	<i>busc-o</i>
2 <sup>nd</sup> SG	<i>busc-as</i>
3 <sup>rd</sup> SG	<i>busc-a</i>
1 <sup>st</sup> PL	<i>busc-amos</i>
2 <sup>nd</sup> PL	<i>busc-aís</i>
3 <sup>rd</sup> PL	<i>busc-an</i>

In the first example, I illustrate the resulting difference with regards to the number of morphemes and to generalisation. The second example below illustrates how the frequency of occurrence of individual morphemes differs between the two systems.

The numerical divide between morphemes in agglutinating and fusional systems is reflected in Plank (1986). He notes that the number of exponents necessary to express the entire paradigm in a system using separative exponence is calculated by adding up all values of all features. In contrast, the number of required exponents in a system with cumulative exponence is calculated by multiplying all values of all features which are expressed cumulatively. Plank (1986) illustrates this using a fictive paradigm with four-term CASE and two-term NUMBER, in which altogether eight complex meanings can be expressed. **Figure 3** shows how six individual separative feature values can express these eight meanings, compared to **Figure 4**, which shows how eight cumulative forms would be necessary to express the same, since every combination of meanings is expressed holistically by a different morpheme. Note that this means that in principle, once a learner has observed the six compositional signals N1, N2, C1, C2, C3 and C4, they can construct all eight complex meanings, if the order of the features following the noun stem is understood. In other words, a learner can generalise the full system once all values have been observed (Brighton, 2002). In a holistic system (**Figure 4**), every combination of meanings needs to be observed individually through fusional morphemes until all possibilities of complex meanings have occurred in the input.

Noun-N1-C1	Noun-N2-C1
Noun-N1-C2	Noun-N2-C2
Noun-N1-C3	Noun-N2-C3
Noun-N1-C4	Noun-N2-C4

**Figure 3:** Paradigm expressing NUMBER-CASE combinations through separate exponents. Note that this system consists of six distinct exponents: N1, N2, C1, C2, C3, C4. (*adapted from Plank, 1986*)

Noun-N1C1	Noun-N2C1
Noun-N1C2	Noun-N2C2
Noun-N1C3	Noun-N2C3
Noun-N1C4	Noun-N2C4

**Figure 4:** Paradigm expressing NUMBER-CASE combinations through eight cumulative exponents. The noun stem is followed by a single cumulative exponent in each case. (*adapted from Plank, 1986*)

Every additional feature and every additional value per feature in a paradigm increases this numerical discrepancy between compositional and holistic systems, since new feature values are observed as fast or faster than new holistic morphemes (Brighton, 2002). Further, “[a]s more objects are observed, the likelihood of full feature value coverage increases rapidly in comparison to the degree of object coverage. [...] Expressivity is a function of object coverage for holistic language. Expressivity is a function of feature value coverage for compositional language (Brighton, 2002, pp.47-48).” This is why a learner can likely express more meanings after the same amount of exposure to the language when acquiring a compositional compared to a fusional system. Thus, the first numerical advantage of compositional systems is that they contain fewer forms to be learned, which impacts on generalisability.

Patterns typical of inflectional systems in natural languages further increase this theoretical difference in numbers. Plank (1986) observes that it is typical for paradigms with separate exponents to have unmarked members in their categories. For instance, this occurs with singular number and nominative case in Turkish (Comrie, 1981). If this is the case, the number of overt separate forms in **Figure 5** might reduce to only four. On the other hand, fusional systems typically contain additional formal idiosyncracies, which increase the number of different forms even further (Plank, 1986). Examples of this are forms reflecting phonological assimilation to different stems or inflectional classes. Note that in natural language paradigms,

syncretism of forms may occur as well, presenting a counter force to the increase in affix variation. For instance, in Russian, the nominative plural and the accusative plural suffix is the same (-y, as in *stol-y*, ‘tables’) in declension Ia and II, but Russian affixes can in turn vary depending on declension class (Comrie, 1983, see example (2) in **Section 1.1**). Overall, one may expect agglutinating systems to be able to express more formal distinctions due to their more economical use of morphemes (Plank, 1986). In short, agglutinating and fusional techniques differ in their number of morphemes as soon as they express more than two binary features. In this special instance, there are four different feature values (such as animate, inanimate, singular, plural), which can yield four complex meanings which are expressed by either four agglutinating or four fusional morphemes (animate\_singular, animate\_plural, inanimate\_singular, inanimate\_plural). However, even in this special case, additional characteristics that often go in hand with both morphological techniques slightly enhance the numerical divide between agglutinating and fusional paradigms.

Given the examples above, we are able to see why transparent morphemes occur more frequently in a learner’s input. By default, a prototypical agglutinating system concatenates meanings. When considering the example with binary features, the form for animate will occur twice: in animate\_singular and in animate\_plural. Analogously, the same holds true for each of the four feature values. In a fusional system, the four different forms do not necessarily share any such formal overlap, although they share elements of meaning.

As Goldschneider & DeKeyser (2001) observe, “[t]hird person -s conflates person, number, tense, and aspect. Only when these four independent components of meaning happen to coincide does this morpheme (or more precisely, morph) make its appearance” (p.36) (see also DeKeyser, 2005). This has an important implication, namely that a learner of a prototypical agglutinating system will see each morpheme more often than a learner of a fusional system due to the frequent re-combination of morphemes with other morphemes. Another way to view this is that a learner of a compositional system will cycle through all possible forms in the language more quickly and more often than a learner of a holistic system, which leads to a difference in exposure to each morpheme. Given that an agglutinating paradigm requires fewer distinct morphemes to express the same meaning space than a fusional paradigm, this may also reduce the strain on memory when acquiring the system (see e.g., Anttila, 1989; Plank, 1986). In short, compositional agglutinating systems typically

require fewer morphemes, which in turn repeat more frequently than fusional morphemes.

### **2.3.2 Implications of the higher frequency of morphemes for learnability**

The effects of frequency on the acquisition of first and second languages have been discussed widely (Larsen-Freeman, 2002). The straightforward hypothesis hereby is that morphemes which occur more frequently in the input will be acquired earlier (Kwon, 2005). Indeed, a range of research studies have shown that higher frequency of morphemes in the input aids learning (Ellis & Collins, 2009; Gass & Selinker, 2008; Goldschneider & DeKeyser, 2001; Larsen-Freeman, 2010; Pfänder et al., 2013; Xanthos et. al., 2011). During the morpheme order studies, frequency was handled as a potential predictor of learning ease by several authors (e.g., Andersen, 1978; Brown, 1976; de Villiers & de Villiers, 1973; Larsen-Freeman, 1976). Goldschneider & DeKeyser (2001) also describe input frequency as the second most proposed predictor of morpheme acquisition order and they include it as one of the predictors in their model. They find that frequency in combination with other factors (among them transparency and regularity) explains a large part of the order of acquisition of English grammatical morphemes in L2 learning data. While some studies have not been able to demonstrate a significant effect of frequency on the learnability of morphemes (e.g. Brown, 1976; de Villiers & de Villiers, 1973), others proposed that it is one of the crucial predictors. For instance, Larsen-Freeman (1976) tentatively suggests that “morpheme frequency of occurrence in native speaker speech is the principle determinant for the oral production morpheme order of second language learners”, having observed significant effects of frequency in the input for L2 acquisition of English (p.132). An effect of frequency is also confirmed in Andersen (1978). At the same time, Larsen-Freeman and others assumed in the 1970s that frequency alone could not explain all of the acquisition difficulty observed at the time (see Andersen, 1978; Larsen-Freeman, 2002).

Nevertheless, frequency has implications for language learning and processing in manifold ways. The segmentation and acquisition of morphological markers in the input to children is assumed to be affected by frequency as well, in that the more frequent markers and those appearing with a larger variety of stems are learned earlier than more infrequent ones (Penke, 2012). Research has also suggested that higher

type frequency of a marker, that is the number of different stems that a marker occurs with, leads to earlier identification of the marker as an invariant form in the input to children, which also impacts on productivity (Bybee, 1995; Penke, 2012). Collins et al. (2009) find that type frequency increases the overall salience of morphosyntactic markers in L2 input. Recognition of a morphological marker is a necessary first step to understanding the meaning it conveys. Penke (2012) connects these claims with findings by Bittner et al. (2003) which show that the type of inflectional system also plays a role: children acquiring more transparent and morphologically richer systems are faster in detecting and acquiring miniparadigms than when their language is less transparent. Frequency may play a role here in the sense that non-transparent systems may display a larger number of distinct morphemes, which increase the number of different forms in a paradigm, and they may display inflectional classes, where a certain marker only occurs with certain stems. On the contrary, agglutinating systems typically display fewer inflectional classes and less stem and affix variation, which implies that individual morphemes occur more frequently.

Thus, frequency of occurrence may be crucial in order to notice morphemes as units in the input, which has been suggested to impact processing. Rastle & Davis (2008) describe the possibility that learners observe regularities in the orthography of a language and keep track of highly frequent letter sequences as morphemic units. For instance, the suffix *-able* can attach to a range of different stems, and although it is not the most frequent morpheme in English, this variation in stems enforces the suffix as an individual unit. In this way, frequency might impact how morphology is processed, but it may also influence which units are noticed first. Additional research has claimed that morphological richness and regularity impact favourably on learning ease, and it is plausible that the frequency of individual morphemes is tightly connected to these (Laaha & Gillis, 2007; Slobin, 1973).

## **2.4 Summary of the effect of transparency and related predictors**

As **Section 2.1** has outlined, transparent structures have often been understood as more learnable than non-transparent ones. However, it is important to note that transparency in morphological paradigms in natural languages often goes in hand with additional features. First, transparent separative exponents are by default part of a compositional system, as long as they fall under synthetic morphology. As described

in **Section 2.2**, compositionality has been shown to emerge under pressures for learnability in numerous studies using iterated learning designs, suggesting that it supports acquisition. Compositionality itself also has repercussions in a number of areas. For instance, it allows for generalisation, which is advantageous because it allows the learner to acquire the language faster and with less exposure compared to a fusional language. Having observed all the feature values of the language, a learner can generalise them to previously unseen complex meanings. Third, **Section 2.3** has illustrated that compositional systems have a numerical advantage because they require fewer morphemes, which in turn repeat more frequently due to their recombination in expressing complex meanings. This connectivity between characteristics of agglutinating systems is important because it needs to be accounted for when assessing the impact of transparency.

## **2.5 The role of the mother tongue and transfer in the acquisition of L2 morphology**

In the previous sections I have illustrated several characteristics pertinent to morphological structures which should be considered when making claims about their impact on learnability. Further, the artificial language learning experiments I employ throughout this thesis have similarities with L2 learning situations and it is important to consider how knowledge from the L1 might impact acquisition, especially since it is possible that experience with certain morphological techniques from the mother tongue plays a role in L2 acquisition. After all, one could imagine that experience with synthetic morphology in general will be helpful when acquiring another synthetic language. More specifically, it might also be the case that native speakers of a relatively transparent language will have a particularly strong preference for transparent structures also in second languages. The following sections will introduce the familiar question of language transfer and the question whether abstract categories can be transferred as well, morphological technique potentially being one of them. Finally, we also need to consider the likely differential processing of inflectional morphology between L1 and L2 speakers to ensure that L2 learners, as tested in the experiments presented in this thesis, perceive morphs and morphemes, an implicit assumption which claims about the beneficial role of transparency entail. In

fact, research suggests that L2 processing strategies may be influenced by the morphology in the L1. The subsequent sections will introduce these topics in brief, since a more detailed discussion of the potential role of transfer and influence from the mother tongue will be provided in **Chapter 4**.

### **2.5.1 Native language transfer in L2 learning**

The possible role of native language transfer has been of interest since the time of the ancient Greeks, but research on transfer has increased considerably over the past decades with the inception of second language acquisition as a linguistic sub-discipline (Jarvis, 2015). Similar to transparency, L1 transfer has been postulated as an important factor during the morpheme order studies in the 1970s and onwards (Andersen, 1978; see also Fathman, 1975, and Hakuta, 1976, as reported in Kwon, 2005). Although the observed acquisition orders were largely similar across different groups of L1 speakers across studies, some researchers observed an impact of grammatical features in a learner's L1 on the ease of acquisition of these features in the English L2 (see Kwon, 2005). Since the 1970s, transfer has been considered in many studies until the present day (Larsen-Freeman, 2002). One area of investigation is transfer of abstract categories from the L1. Abstract features like articles or grammatical gender have been found to pose difficulties when they are not familiar from the L1, partly due to their abstractness (DeKeyser, 2005). Research also suggests that familiarity with morphological structures from the L1 facilitates acquisition of L2 morphology because it increases sensitivity to morphemic structures. Some examples of these different kinds of transfer will be highlighted in the next sections.

### **2.5.2 Transfer of abstract categories from the L1**

A range of abstract categories have been considered under the aspect of transfer from the L1, the assumption being that categories which are familiar from the L1 will be beneficial in the acquisition of an L2. For example, Portin et al. (2007) suggest that the lack of features like GENDER and DEFINITENESS in Finnish could increase the offline processing cost of L2 Swedish nouns inflected for these features, since neither conceptual nor translation equivalence exists in the L1. The authors propose that especially because the DEFINITENESS suffix depended on the GENDER of the noun in

their study, lack of experience with GENDER in the L1 might have resulted in long reaction times for low-frequency nouns. Relatedly, a study by Sabourin et al. (2006) found that the acquisition of grammatical GENDER marking in an L2 is facilitated when GENDER is marked in a morphologically similar way in the L1.

However, difficulties with novel concepts may not only be explained by their abstractness alone. As Gor (2010) summarises, research also suggests that the redundancy of certain grammatical markers contributes to acquisition difficulty. Features such as GENDER agreement or plurality are typically already expressed elsewhere in the sentence, which may explain omission errors in some learners (Jiang, 2007, as referred to in Gor, 2010). This consideration of informativeness may also help explain why some features have been found to be difficult for learners even when present in their L1, and may explain some of the difficulty with the early Contrastive Analysis Hypothesis, according to which only those features should pose difficulties which are different between L1 and L2 (Gass & Selinker, 2008). Assuming that morphological technique also represents an abstract category that can be transferred onto an L2, the question of interest for the present study is twofold. First, we need to ask whether knowledge of a certain morphological technique from the mother tongue will improve learning of the same technique in an L2. Second, it remains to be seen whether certain techniques are easier to learn in general, independent of experience through one's native language. These questions will be addressed throughout **Chapters 3-5**.

### **2.5.3 Transfer effects arising from the morphological structure in the L1**

A related question to the one about transfer of abstract categories is whether the morphological system present in the L2 benefits from positive transfer of the morphological system in the L1. At least in a more indirect way, transfer effects of some sort appear plausible. Indeed, experience with decomposition through the L1 has been found to be more helpful for the acquisition of GENDER marking in a novel language than familiarity with the grammatical feature from the L1. Havas et al. (2015) compared performance between L1 Finnish speakers and Spanish-Catalan bilinguals and found that Finnish speakers were more successful in a word recognition and a generalisation task with an artificial language despite their initial unawareness of the feature of grammatical GENDER. These findings suggest that familiarity with

morphological decomposition can benefit learning of a second language when it requires the identification of affixes in morphologically complex words.

A similar positive transfer effect of inflectional structures from the L1 to the L2 is observed in a study by Kaivapalu & Martin (2007). Estonian and Russian learners of Finnish were better at acquiring bound inflections in their L2 that were phonologically, morphologically and semantically similar to their L1, but the extent of this positive transfer also depended on their level of L2 proficiency. Higher proficiency thereby lead to increased positive transfer. Relatedly, Kempe & Brooks (2008) note that experience with grammatical structures which are similar between L1 and L2, such as transparent GENDER marking, can tune learners' attention towards these features and help learn them more easily.

These examples show that transfer effects can be observed for highly specific structures. On a more general level, experience with morphological techniques may cause a different type of transfer effect, namely in the area of online processing as described above. It is possible that extensive morphological processing in the L1 helps learners identify morphemic units in the L2 input. In this case, a morphologically rich and transparent first language would generally be expected to be more facilitative than an inflectionally poorer or a more opaque language, since it attunes the L2 learner to identifying morpheme units in the input, independent of the morphology dominant in the L2. Certainly, a language can exhibit more than one morphological technique, as illustrated in **Chapter 1**. However, we can begin to consider possible beneficial effects by looking at first languages that are predominantly using one technique. To this end, I selected participants with L1 English to represent the isolating type, L1 Turkish to represent the agglutinating type and L1 Spanish to represent the fusional type. **Chapters 3-5** will present the findings for each speaker group in turn.

#### **2.5.4 The differential processing of morphology in L1 speakers and L2 learners**

Independent of experience with abstract categories or morphological techniques from the mother tongue (or potentially even from another L2), one should also ask the question to what extent L2 learners are sensitive to the morphological level in a second language in the first place. While learners will acquire L2 morphology at least to some extent, this area has been found to pose difficulties for learners (Clahsen et al., 2010;

Prévost & White, 2000). Since processing (online and offline) arguably underlies language acquisition, sensitivity to morphemic units should conceivably enhance learning ease, while a lack of such sensitivity might lead to learning difficulty. Especially since the experiments presented in this thesis test the earliest stage of acquiring a second language, and since the focus of the present research is on the supposed benefit of transparent morphemes, which rests on morpheme segmentation, it is important to consider whether learners are indeed sensitive to this level at this early stage. To emphasise, this sensitivity is essential in order to exploit the benefits of agglutinating systems (transparency, compositionality, generalisability, frequency), although it is often left implicit in claims in the acquisition literature. If a learner does not segment words into their morphemes, they inevitably need to process and memorise larger chunks such as *oynu-yorsunuz* versus *spiel-t*, in which case no learning advantage of an agglutinating system is to be expected. However, it has often been stated that L2 learners struggle with morphology and that they fail to produce obligatory morphological forms (DeKeyser, 2005; Kimppa et al., 2019).

Research on the processing of morphologically complex words has often described L1 and L2 processing as different (Clahsen, 2010; Gor, 2010). Mainly two routes of processing are generally being considered: a speaker or learner of a language can process a morphologically complex word by decomposition into its constituent morphemes or complex words can be stored and accessed as whole-word representations. Using the former route means to exploit the words' compositional structure and saves storage space as far as long-term memory is concerned. Full-form representations, on the other hand, are faster, but demand more storage space (Portin et al., 2007). It is generally assumed that speakers use both routes, but the extent to which this happens in L1 and L2 processing is still debated, and likely differs between online and offline processing.

In the domain of L1 online processing, segmentation of morphemic units is often described as early and automatic (Lewis et al., 2011; Rastle & Davis, 2008; Zweig & Pylkkänen, 2009). However, proposals for a dual-route mechanism have also been made, according to which full-form activation and morpheme-based processing happen simultaneously, a process which ultimately speeds up processing (Baayen et al., 1997; Gor, 2010). In comparison, less is generally known about online morphological processing in L2 learners and the exact extent to which both possible processing routes are activated in L2 speakers is debated (Kimppa et al., 2019). While

some have claimed that even advanced learners rely more on declarative than on procedural memory (Clahsen et al., 2010), the opposite claim has been made as well, namely that whole-word storage occurs only with very high-frequency words, while less frequent words are decomposed (Portin et al., 2007). Overall, L2 learners appear to be slower in online processing than L1 speakers, which signals that processing is costly (Gor, 2010; Kimppa et al., 2019). The exact nature of online processing employed by L2 speakers, however, is still underexplored.

Regarding offline processing, the differences between L1 and L2 speakers seem less pronounced and some studies have found that L2 learners use decomposition in behavioural tasks (Bosch et al., 2017; Portin et al., 2007). Additionally, studies have suggested the potential for a development towards L1-like processing: with increased exposure and proficiency in an L2, learners may become more similar to native speakers in the use of their processing strategies, a finding which has been made both online and offline (Bosch et al., 2017; Kimppa et al., 2019; Portin et al., 2007). However, the experiments presented in this thesis test the earliest stage of L2 learning and again, less is known about processing in this context. However, whether offline segmentation of morpheme units happens during the experimental procedure has been tested as part of the studies presented in this thesis, and will be reported in **Chapter 3**.

Finally, it is also conceivable that morphological richness in the L1 and exposure to an L2 influence processing in both languages. After all, it has been found that in Finnish, a morphologically rich language, medium-frequency inflected nouns are processed via decomposition by monolinguals, while inflected words of a similar medium-frequency range in Swedish were processed via the full-form route by monolinguals (Portin et al., 2007). These results indicate that the cut-off point between decomposition (typical of lower frequency words) and full-form access (typical of high frequency words) may be modulated by the morphological richness of a language. However, a study by Lehtonen et al. (2006) also found that Finnish-Swedish early bilinguals used decomposition of low and medium-frequency Swedish words, which stands in contrast with Swedish monolinguals, who decompose only low-frequency forms. In a similar vein, inflectionally rich L1s like Finnish and Hungarian have been shown to increase the amount of decomposition in L2s, even beyond the level found in native speakers (De Diego-Balaguer, 2010, referring to Portin et al., 2008; Portin & Laine, 2001). These findings could suggest that apart from the morphological richness

of a language, also the amount of exposure to a language plays a role. Since bilinguals and L2 speakers receive less input from either language, individual words may have a lower frequency count for them in general, resulting in more decomposition. With regards to the experiments presented in this thesis, this could mean that at the earliest stage of acquiring a new language, words have been observed with very low frequency, which could promote decomposition, and therefore enable the learner to profit from compositional and transparent structure. I will return to this point in **Chapter 4**.

## **2.6 Previous L2 research on implicit learning of artificial grammars**

Before concluding the chapter and moving on to the language learning experiments, a final aspect deserves attention which comes from previous research on implicit learning. Since the experiments that will be presented in this thesis require participants to learn grammatical categories and their mapping onto grammatical morphemes under implicit learning conditions, it is insightful to see how adult participants in other studies fared with similar tasks and grammatical features. Implicit learning can be described as “learning that proceeds without awareness of what is being learned and without intention to learn it” (Leung & Williams, 2011, p.33). This means that participants are not told about the structures they are required to learn and therefore have no intention to learn them. Secondly, implicit learning is usually taken to refer to situations in which learners are not able to verbalise their acquired knowledge because it is unconscious. Nevertheless, their performance in the learning task suggests that they did acquire the new system (Rebuschat & Williams, 2012).

Previous research has found that learners are able to acquire regularities in their linguistic input without actively paying attention in word segmentation or statistical learning tasks (Leung & Williams, 2012). The focus of investigation has been further expanded to test whether form-meaning relationships can also be implicitly learned in artificial grammar learning experiments. More specifically, the question at hand is whether participants are able to form an association between a word or morpheme and a concept (for instance activated through an image) which they are given as stimuli, without explicitly being told or made aware of the underlying connection (see Leung & Williams, 2011).

A study by Leung & Williams (2012) finds that adult learners are able to acquire mappings between forms (morphemes) and grammatical meanings relating to ANIMACY, but not to object size. In their first experiment, L1 English-speaking participants learned a correspondence between determiners and nouns which was based on the ANIMACY of the noun. Since participants were not able to learn a similar relationship when it was dependent on the size of an object, their findings suggest that ANIMACY as a feature can be learned in experiments under implicit learning conditions (Leung & Williams, 2012). This is important to highlight because the feature of ANIMACY is also inbuilt into the artificial languages used in this thesis.

Participants in Leung & Williams (2012) were introduced to four determiners, *gi*, *ul*, *ro*, and *ne*, whereby they were told that *gi* and *ul* were used with near objects and *ro* and *ne* were used with far objects. However, there was another underlying distinction which participants were not told about, namely that *gi* and *ro* referred to animates and *ul* and *ne* referred to inanimates. For instance, they heard statements such as *gi goldfish* or *ro bull* while seeing two images. This means that each of the four determiners was used for a unique combination of two values such as *gi* meaning near+animate. In a generalisation task with novel nouns and images, participants were exposed to grammatical and ungrammatical determiner-noun combinations and had to make a decision about the object in the picture by clicking the mouse. Reaction times of those participants who reported to have noticed the relationship between the articles and the ANIMACY distinction indicated that they were able to decide which picture a statement referred to based on the articles alone, before the noun information became available. However, even those participants who did not report to have noticed a correlation with ANIMACY values produced slower reaction times in trials where the association was violated, which indicates that they implicitly learned the pattern.

Leung & Williams (2012) conducted a further experiment in which participants were presented with the same stimuli, but were told that determiners were conditioned on ANIMACY, and were not told that determiners were also conditioned on intrinsic object size. In this experiment, only three out of 26 participants were able to learn and use the correspondence between determiners and intrinsic object size alone during the generalisation task. Response times did not slow down during trials in which the correspondence was violated, which suggests that the underlying association was generally not learned. This difference between both experiments suggests that

humans readily classify items based on an ANIMACY scale, but not based on item size, which links to ANIMACY playing a prominent role in language grammars (Vihman & Nelson, 2019).

In a similar experiment to the previous ones, Leung & Williams (2011) introduced adult participants to the same four grammatical morphemes which this time functioned as articles. Participants were told that one pair (*gi*, *ro*) was used with personal names of adults and the other pair (*ul*, *ne*) was used with personal names of children. The hidden dimension was that of the thematic role of the noun phrase: agents required the articles *gi* and *ul* while patients corresponded with *ro* and *ne*. In a reaction time test, participants were exposed to pictures and accompanying English-like auditory sentences including the novel articles (such as *Kiss ul Mary a boy on the face*) and needed to perform a decision task about the position of the named person in the picture. Although 20 out of 25 participants reported to not be aware of the mapping between articles and thematic roles, their reaction times were significantly shorter in the grammatical compared to the ungrammatical test trials. This study therefore extends research findings on implicit learning of form-meaning mappings to involve thematic roles, which can only be accessed when contextual information is taken into account (in contrast to accessing the meaning of individual words) (Leung & Williams, 2011). Again, the results suggest that participants are able to associate morphemes in an artificial language learning task with meaning dimensions which they are not explicitly told about.

A final study by Marsden et al. (2013) shows that in early stages of L2 acquisition, learners develop sensitivity to morphemic units and their associated meanings. They present results from three experiments based on the same miniature artificial language that consisted of stem-suffix combinations which were presented both orally and visually through images. Using a crossmodal priming task, they test the conditions under which grammatical suffixes are detected and segmented from the stem, whether participants' attention needs to be oriented to the suffixes to do this, and what role is played by the attribution of meaning to stems and suffixes.

They find that when no meaning is involved in learning (Experiment 1), participants become familiar to the word structure and the surface forms of the suffixes during exposure, but do not form abstract morphological representations of the suffixes, since no priming effects were observed. When the task involved explicit attribution of meaning to stems and implicit correlations between suffixes and

grammatical meanings (Experiment 2), Marsden et al.'s (2013) results suggest that participants had acquired the underlying relationship between suffixes and their meanings, although they mainly reported to be guessing or selecting answers intuitively. This lack of awareness of rules suggests implicit learning even when participants' attention was not directed to the suffixes, but to the stems. When explicit information was provided about the forms and meanings of the suffixes during exposure (Experiment 3), suffixes were learned considerably well. Additionally, participants reported explicit knowledge of the suffix-meaning relationships to a greater extent than in the previous experiments, and when they used this explicit knowledge to inform their decisions in the task, it led to higher accuracy compared to using implicit knowledge. Participants further showed incidental learning of the stems in this experiment. However, none of the experiments in Marsden et al. (2013) showed any crossmodal priming effects, which suggests that abstract underlying representations of suffixes are not formed under these learning conditions.

This study suggests that purely orthography-based decomposition of stems and suffixes can already occur at initial stages of learning (in this case induced by the use of an artificial language) and even in the absence of meaning, since learners show sensitivity to the suffixes. Further, Marsden et al. (2013) show that the meaning of suffixes can be learned implicitly even when participants' attention is not directed to the grammatical part of the word. The findings therefore add to those from the studies conducted by Leung and Williams (2011, 2012) discussed above. These results are relevant for the experiments to be presented in the next chapters, where learning form-meaning mappings is also measured at the very first stage of acquiring novel morphology and where participants are not explicitly informed about the meaning of suffixes.

To summarise, results from the studies discussed in this section show that adult participants are able to acquire grammatical form-meaning mappings in artificial language learning experiments even when these mappings are not explicitly brought to the participants' attention. They further suggest that based on the orthography of words, learners are able to detect suffixes as morphemic units. Finally, results from these experiments suggest that ANIMACY as a grammatical feature is accessible under implicit learning conditions. This is important to register, since the experiments that will be presented in this thesis also involve ANIMACY as a grammatical dimension in the artificial language learning task.

## Summary

This chapter has described the common assumption in language acquisition research that structures with a transparent form-meaning mapping are easier for L1 as well as L2 learners. This belief rests on the notion that when one form corresponds to one meaning and one meaning corresponds to one form, this is easier for a learner than when there is a one-to-many or a many-to-one relationship between meaning and form, as is the case with cumulation, multiple exponence or syncretism. We also noted that transparent forms are a hallmark of agglutinating systems, and non-transparent forms are typical of fusional systems, which has led some scholars to claim that agglutinating structures are overall more learnable than fusional ones (Hengeveld & Leufkens, 2018; Igartua, 2015). Notably, transparency in natural synthetic languages coincides with additional factors: transparent morphemes occur in compositional systems, since agglutinating morphemes are concatenated into longer words. Compositionality comes with further characteristics of its own; morphemes occur more frequently than they do in holistic systems, which is due to the compositional nature, but also the regularity of these systems (lack of allomorphy, lack of inflectional classes, etc.). A second consequence of compositionality is that a learner can generalise the language to novel complex meanings without having observed the corresponding signals in the input. These additional factors need to be taken into account when assessing claims about the beneficial role of transparency for language learning. Finally, also the experience from the mother tongue in terms of morphological technique, the possibility of language transfer and the role of morphological processing need to be considered when testing the effect of transparent form-meaning mappings. These factors will be tested in **Chapters 3-5**, where the comparison of different L1 speaker groups allows for an investigation of different constellations of morphology present in the L1 and in the L2.

## 2.7 Research questions

Having outlined the research context on the acquisition of inflectional morphology and several factors that likely play a role as well as the difficulty of demonstrating a beneficial effect of transparency alone on learning, several research questions arise. First, is it possible to trace learnability back to transparency in and of itself, when using

an artificial language-learning paradigm that allows us to keep other factors constant? Second, are transparent structures easier to learn for L2 learners independent of their differing experience with transparent morphology through the mother tongue? Or does the morphological make-up of the mother tongue affect what type of morphology will be easy to learn in an L2? Lastly, can we be certain that L2 learners are sensitive to the morphological level, especially in the early stage of acquiring an L2? The role of transparency for learning and the success of segmentation will be tested elaborately in **Chapter 3**. The role of the type of morphology present in the L1 will be tested in **Chapters 4** and **5**. A comprehensive answer to all of the questions raised here will be provided in **Chapter 6**.

## Chapter 3

### Testing the effect of transparency in artificial language learning in native speakers of English

#### 3.1 Preface

One of the main observations in **Chapters 1** and **2** was that natural language paradigms exhibit multiple characteristics, such as different levels of regularity, compositionality, and different frequencies of individual morphemes. Assessing only the effect of the transparency of morphemes on learnability in natural languages therefore becomes very difficult. For this reason, I devised an artificial language learning paradigm which allows me to reduce the influence of the number of morphemes and their regularity in order to solely manipulate the factor of interest, namely the level of transparency of inflectional morphemes. Artificial language learning experiments are a common method in linguistics for this purpose (Culbertson, 2012; Culbertson & Schuler, 2019; Levshina, 2019).

As stated earlier, I tested the impact of different transparency levels on the acquisition of inflectional morphology in three different L1 speaker populations. **Chapter 3** details the results of experiments run with the L1 English speaker group. This is in the form of a submitted paper which is currently under review (**Section 3.2**). It is co-authored by myself and my supervisors Dr. Jennifer Culbertson and Prof. Kenny Smith. I began by testing native speakers of English because this provides a relatively neutral testing ground: English is typically described as a largely isolating language, having lost most of its inflections between the Old English period and its present-day state (Kortmann, 2005). Indeed, English only has 8 inflections, namely plural *-s* and genitive *-s* on nouns, third person singular present tense *-s*, present participle *-ing*, past *-ed*, past participle *-ed* on verbs, and comparative *-er* as well as superlative *-est* on adjectives (Kortmann, 2005). Therefore, I assume that native speakers of English are neither highly familiar with agglutinating morphology nor with fusional morphology. This means that we can rule out native language transfer of inflectional structures as a factor contributing to learning ease in either of the two conditions tested. If we observe a learning advantage in one condition, we are able to attribute it to transparency, which is the factor that differentiates both conditions. The

experimental paradigm used in **Chapter 3** is largely the same as used in the remaining experiments in this thesis.

### **3.2 The role of transparency in the acquisition of L2 morphology: artificial language learning experiments with native speakers of English**

#### **Abstract**

Agglutinating morphology has often been argued to be easier to learn than fusional morphology. One potential explanation for this is that agglutinating structures involve a one-to-one mapping between meaning and form, and are thus more transparent than fusional ones, where a single morpheme expresses multiple meanings. However, it is difficult to demonstrate a benefit of transparency in natural language learning, where multiple potentially confounding factors are at play. Here, we investigate the role of transparency in the learnability of inflectional systems under tightly controlled conditions in an artificial language learning paradigm. Contrary to our predictions, we find that English-speaking learners do not find a more transparent agglutinating system to be more learnable than a less transparent fusional system. We further explore two additional common differences between agglutinating and fusional systems: the number of morpheme boundaries per word, and word length – both of which are typically greater for agglutinating systems. Controlling for these differences fails to reveal any advantage for agglutinating systems. Our results suggest that the advantages of agglutinating systems may lie elsewhere, e.g. in the generalisability advantage arising from the compositional nature of agglutinating systems, or from a combination of factors of which transparency is only one (and perhaps quite minor), features which may only come into play when the morphological system is relatively large.

**Keywords:** language acquisition, morphology, agglutinating, fusional, transparency, artificial language learning

#### **1. Introduction**

##### **1.1 Agglutinating and fusional morphology**

Languages differ in the ways they express grammatical meanings morphologically. While some languages, like English, use relatively little inflectional morphology, other languages, such as Turkish and Russian, have much richer inflectional systems. Such

languages may use morphemes to express features like number, gender, case, etc., which are added to the stem to form words. Importantly, among languages that use inflections to convey grammatical meaning (i.e., synthetic languages), one can make a further distinction between languages using *agglutinating* morphology and those using *fusional* morphology.

While not all agglutinating or fusional systems display exactly the features common to that morphological type, and not all to the same extent, perhaps the central feature that distinguishes these types is exponence, which relates to how grammatical meanings are mapped onto individual morphemes (Igartua, 2015; Plungian, 2008; Bickel & Nichols, 2013).<sup>4</sup> Agglutination is characterised by separative exponence of grammatical meanings – each morpheme typically only carries one meaning and they are combined linearly to form words. Fusion, on the other hand, is characterised mainly by cumulative exponence – a single grammatical morpheme expresses more than one grammatical meaning. Multiple fusional affixes can also be combined linearly in a word (Igartua, 2015; Plank, 1986; Plungian, 2008; Bickel & Nichols, 2013). Example (1a) illustrates cumulative exponence: the fusional suffix *-t* expresses three grammatical meanings at once (present tense, second person, plural number). This stands in contrast to (1b) where three separate agglutinating suffixes (*-yor*, *-sun*, and *-uz*) individually convey the same pieces of information. The distinction between separative and cumulative exponence, and its implications for learning, are central to this paper, as discussed further below.

- |                               |  |
|-------------------------------|--|
| 1. (a) German (fusional)      | (b) Turkish (agglutinating)                    |
| <i>spiel</i> <i>-t</i>        | <i>oynu</i> <i>-yor</i> <i>-sun</i> <i>-uz</i> |
| play 2 <sup>nd</sup> .pl.pres | play pres 2 <sup>nd</sup> pl                   |
| ‘You play.’                   | ‘You play, you are playing.’                   |

In natural languages, agglutinating and fusional systems go hand in hand with further characteristics. For instance, fusional systems typically display more affix and stem alternations (i.e., allomorphy), inflectional classes and syncretisms relative to agglutinating systems. In addition, fusional affixes typically cannot be attached to different categories of words, while agglutinating affixes often can, and the linear order

---

<sup>4</sup> In addition, while some languages can be characterised as predominantly agglutinating or fusional, these two strategies can also both be found within a single language (Plungian, 2008; Haspelmath, 2009; Igartua, 2015).

of morphemes in a word tends to be more restricted in fusional than in agglutinating words (Haspelmath, 2009; Plank, 1986; Plungian, 2008). These features mean that agglutinating systems typically display more regularity in how forms correspond to meanings. Two additional differences are worth mentioning here. First, agglutinating systems have been claimed to form larger morphological paradigms than fusional systems (Plank, 1986). Second, agglutinating words tend to be longer than fusional words (Dressler, 2005), likely because agglutinating paradigms can express more combinations of grammatical meanings through compositional use of (fewer) morphemes. Longer word length is a direct consequence of the separative exponence of agglutinating morphemes – the same number of grammatical meanings would be expressed in a single, typically shorter fusional morpheme (Plank, 1986; Plungian, 2008). This is illustrated in the German and Turkish examples in (1a) and (1b) above.

## **1.2 Hypothesized learnability advantages of agglutinating systems**

Agglutinating morphology has long been argued to have a learnability advantage over fusional morphology. Building on the differences identified above between these two types of systems, there are two main factors that have been proposed in the literature to lead to better learnability of agglutinating over fusional systems: transparency and compositionality.

### **Transparency**

A main general difference between agglutinating and fusional systems that has been claimed to affect learnability is transparency. In the literature, this comprises two slightly different, but related, notions, both discussed above in contrasting agglutinating and fusional systems. First, a one-to-one correspondence of individual form-meaning pairs, i.e., separative exponence. We refer to this here as transparency and distinguish it from what we will call regularity, which operates on the system level, and is characterised by the absence of variation in affixes (allomorphs), syncretism, inflectional classes, etc.

In a large typological survey, Bickel & Nichols (2013) find that for the features of case and TAM (tense-aspect-mood), separative exponence is more common than cumulative exponence with other features. However, morphological change has often been described as a cycle from isolation (no inflectional morphology) to agglutination to fusion and finally back to isolation (although see Igartua (2015) for cases of change

from fusion to agglutination). Nevertheless, the view from typology and diachronic change seems to be that agglutinating systems are the default, less complex, or in some sense more 'natural' or 'optimal' and thus expected to be easier to learn (Anttila, 1989; Bybee, 1985; Bickel & Nichols, 2013; Don, 2017, p.133; Igartua, 2015, p.709; Hengeveld & Leufkens, 2018). In accord with this, research on both first and second language acquisition of morphology has argued that the learnability of morphological systems depends in part on the level of transparency and regularity (DeKeyser, 2005; Dressler, 2012; Goldschneider & DeKeyser, 2001; Penke, 2012; Slobin, 1973).

A number of studies have suggested advantages for transparency and regularity in first language acquisition. In their classic studies on morpheme acquisition order, Brown (1976) and de Villiers & de Villiers (1973) found that English-learning children take longer to acquire the fusional morpheme -s as in *plays* (3rd person, singular, present tense) than the more transparent and more regular *-ing* (progressive). Slobin (1973) argued that transparency and regularity contribute to early learning of morphology in Hungarian and Turkish compared to e.g., Serbo-Croatian (see also Aksu-Koç and Slobin (1985), Hengeveld & Leufkens (2018) for a comparison to Dutch). Hengeveld & Leufkens (2018) note that this is despite the fact that the Dutch verbal system is simpler than Turkish in terms of the number of morphemes. Narasimhan & Gullberg (2011) report similar findings from a comparison of the acquisition of verbal morphology in Tamil- and Dutch-acquiring children. They target caused posture verbs, which are expressed as two words in Tamil, but conflated in a single morpheme in Dutch verbs. They find that despite a higher frequency of these verbs in Dutch child-directed speech, Tamil-speaking children acquire them more than a full year before Dutch-speaking children. In a longitudinal study of the development of verb morphology across languages, Bittner et al. (2003) find that only two to four months after producing their first verbs, children are able to detect and reconstruct small verbal paradigms if their language is morphologically rich and transparent (Turkish, Finnish, Russian and Croatian). By contrast, children acquiring languages with less transparent verbal systems, including syncretisms and homophony, are shown to take at least twice as long (French, Italian, Yucatec Maya, German, Dutch and English).

Several findings from second language acquisition also support the claim that transparency and regularity affect learning of morphology. For example, Goldschneider & DeKeyser (2001) conducted a meta-analysis of oral production data

of L2 learners' English in order to test the role of five factors on the acquisition of English grammatical morphemes. They found that in combination with frequency, morphophonological regularity, perceptual salience and syntactic category, semantic complexity was a significant predictor of learning ease. Kempe & Brooks (2008) found that L2 learners of Russian fusional case-marking were better able to generalise case inflections when they were presented with examples of transparent structures in their input, which served as cues to inflection class membership.

Taken together, these results suggest that semantic transparency may make learning easier. Nevertheless, in most of these studies transparency is not fully disentangled from other factors which often go hand-in-hand with it, like frequency and regularity. Indeed, in a longitudinal study of a child learning Austrian German, Korecky-Kröll & Dressler (2009) find that frequency in the input is the most important predictor in early acquisition, with productivity, transparency and iconicity only becoming more important in later stages. Laaha & Gillis (2007) make a similar observation; in their study, they argue that form transparency is a by-product of paradigmatic morphological richness in their model, and word transparency is not a significant predictor of the speed of acquisition in children. Finally, the measures used in early studies like Brown (1976) and later by Goldschneider & DeKeyser (2001) cannot distinguish transparency of form-meaning mappings from factors like grammatical complexity or morphophonological regularity at the system-level. Most important for our purposes, the role of separative vs. cumulative exponence as distinct from morphophonological regularity remains unclear.

### **Compositionality**

The distinction between prototypical agglutinating and fusional morphology also has clear parallels to the comparison between compositional and non-compositional or *holistic* structures. Since agglutinating morphology is characterised by separative exponents which are combined linearly to form words, it exhibits compositional structure. In accordance with research by Gottlob Frege (1848–1920), Krifka (2001) defines compositionality as follows: “The meaning of a complex expression is a function of the meanings of its immediate syntactic parts and the way in which they are combined” (p.152). For instance, the verb *walked* is compositional in that the meaning of the word can be inferred from the meaning of its parts, *walk* and past tense

–*ed*. In contrast, the fusional form *taught* is holistic; its meaning cannot be inferred from subparts of the word since the whole signal expresses the whole meaning.

The compositional structure of agglutinating systems has two related implications for learning. First, compositionality leads to generalisability (Brighton, 2002; Brighton et al., 2005). If a language is compositional, the learner can generalise the system to new meaning combinations as soon as the values of all individual features have been observed. For example, in an agglutinating system, in principle a learner need only encounter animate singular and inanimate plural in order to be able to express all four combinations of meaning. In a fusional system, this is not possible; all meaning combinations need to be observed. Second, when comparing compositional to holistic systems, assuming a consistent meaning space, compositional structures repeat affixes more often than holistic structures. To illustrate, if a language expresses NUMBER (e.g., singular vs. plural) and CLASS (e.g., animate vs. inanimate), then in a compositional system, the words for animate singular and animate plural will share the affix for animate. In a holistic system, however, a single affix will be used to express both feature values and thus there is no formal overlap between the two words. If an affix occurs more frequently in the input to learning, then it may be learned faster (Ellis & Collins, 2009; Goldschneider & DeKeyser, 2001; Larsen-Freeman, 2010; Pfänder et al., 2013; Xanthos et al., 2011).

Numerous experimental and computational studies suggest that compositional languages arise and persist under a pressure for learnability (e.g., Chaabouni et al. 2020; Kirby, 2000; Beckner et al., 2017; Kirby et al., 2008). However, Brighton (2002) shows that the advantage of compositional systems over holistic systems is strongest when the meaning space is large, i.e. when it is composed of many features with many values, which maximises the generalisation advantage of compositional systems.

## **Summary**

In summary, typical agglutinating systems exhibit characteristics such as compositional structure and transparent morphemes. Previous studies suggest that transparent morphology may be more easily learned by children and adults (e.g., Brown, 1976; de Villiers & de Villiers, 1973; Goldschneider & DeKeyser, 2001; Korecky-Kröll & Dressler, 2009; Narasimhan & Gullberg, 2011), and that non-transparent systems are sometimes transformed during language acquisition (e.g., Narasimhan & Gullberg, 2011) and language change (e.g., Anttila, 1989; Igartua,

2015) to make them more transparent; evolving artificial communication systems under learning pressures also develop compositional structures which facilitate learning and generalisation (e.g., Beckner et al., 2017; Chaabouni et al., 2020; Kirby, 2000; Kirby et al., 2008; Kirby et al., 2015; Vogt, 2005). Agglutinating systems also allow for earlier generalisability and individual morphemes occur more frequently than in fusional systems, factors which have been found to aid acquisition (e.g., Brighton, 2002; Chaabouni et al., 2020; Larsen-Freeman, 2010; Pfänder et al., 2013; Xanthos et al., 2011). All of these findings have led to the frequent claim that agglutinating structures should generally be easier to learn than fusional ones. However, in studies of natural language, it is very difficult to isolate the effects of different characteristics of interest. For example, regularity, transparency, and generalisability/compositionality often go hand-in-hand, and evidence for the role of transparency alone – i.e., an advantage of one-to-one mappings from grammatical meanings to morphemes – is difficult to find. Here, we use an artificial language learning paradigm to isolate the effect of transparency on acquiring morphological systems while holding other features constant.

## **2. Experiment 1**

In Experiment 1, we test the hypothesis that a difference in transparency of two morphological systems will lead to a difference in learnability. We use a small morphological feature-space marked on common nouns, with only two features, each with two possible values: NUMBER (singular, plural) and CLASS (animate, inanimate). In our agglutinating system, these features are expressed using separative exponence; in our fusional system, these features are expressed by cumulative exponence. This two-by-two space is the only size of meaning space which allows us to distinguish the effect of transparency from an effect of the number of morphemes to be learned (4 in both systems we tested). This is crucial since we know that the number of morphemes affects learning (e.g., Johnson et al., 2020; Plank, 1986). This small space also minimises any advantages of compositionality or frequency that favour agglutinating systems; these advantages would be more pronounced in larger meaning spaces (e.g. with more features or more values per feature; Brighton, 2002). Further, there are no differences in regularity, e.g., no allomorphy, syncretism, etc. between the two systems. In order to observe differences in learnability in this relatively small system,

we use a design which assesses accuracy on every trial rather than only after some period of training. This allows us to track the time-course of learning.

## 2.1 Methods

### Participants

180 participants were recruited online via Amazon Mechanical Turk, all self-reported to be native speakers of English, at least 18 years old, and who possessed a Mechanical Turk qualification indicating they were based in the USA. They were each paid \$4 for their time. Participants were randomly assigned to one of two conditions (94 to the agglutinating condition, 86 to the fusional condition).<sup>5</sup> We used the Mechanical Turk qualification and ID systems to ensure that participants were only able to participate in the experiment once.

### Materials

The artificial language consisted of 96 nouns which referred to either everyday objects such as household items and clothing (inanimates, 48 items) or to animals (animates, 48 items). The language had four suffixes, encoding the binary features NOUN CLASS (animate or inanimate) and NUMBER (singular or plural). These features were chosen because they can be illustrated clearly in images. All stems were monosyllabic and adhered to English phonotactics (e.g., *chas*, *froon*, *jeas*, *mer*, *troas*). Morphemes used in each condition were identical: *-mu*, *-ka*, *-pi*, *-lo*. However, the grammatical meanings they expressed differed between conditions (see **Table 1**). In the agglutinating condition, morphemes each expressed one value of either NUMBER or CLASS. In this condition, the CLASS suffix always preceded the NUMBER suffix. For instance, in **Table 1**, *-mu* expresses animate, *-ka* expresses inanimate, *-pi* expresses singular and *-lo*

---

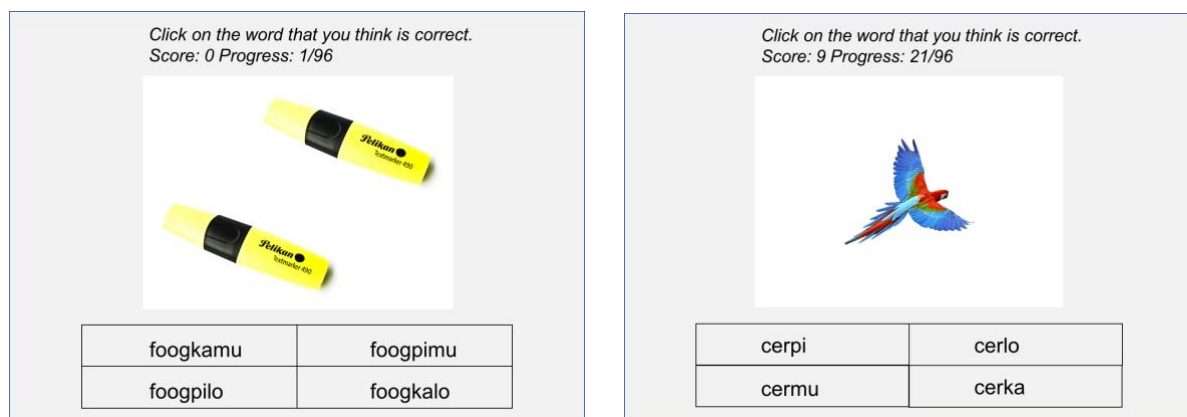
<sup>5</sup> Five additional participants indicated in the questionnaire following the experiment that they were bilingual; their data was excluded. We defined bilingual as speaking more than one language from birth or from a young age, and being fluent in it. Our aim was to exclude participants who had a very high level of proficiency in another language from early in life. 15 additional participants took notes during the experiment, based on their answers in the questionnaire. These participants were also excluded. The remaining data from 180 participants was used for the analysis. Of these, 37 participants indicated having at least basic skills in at least one other language apart from English. Note that the results reported here stem from two separate but identical experiments: An original experiment plus a replication, which we ran to investigate a pattern of error types we found in the initial experiment in an exploratory analysis. We discuss this in detail in **Section 4.4** and in the **General Discussion**. We report the original experiment plus its replication jointly here as a single experiment in the interests of brevity, since the accuracy results by condition are very similar.

expresses plural. In the fusional condition, each of the four morphemes expressed two values, one for each feature. For instance, in **Table 1**, *-mu* expresses animate+singular, *-ka* expresses animate+plural, *-pi* expresses inanimate+singular and *-lo* expresses inanimate+plural. Mappings between morphemes and grammatical meanings were randomised across participants in each condition, but they always adhered to the principles described above.

Note that because we use the same set of morphemes in both conditions, the words are longer in the agglutinating condition (by one 2-letter morpheme). This is a general characteristic of agglutinating languages, where words tend to be longer than in fusional languages as a consequence of separative exponence (Dressler, 2005; Plungian, 2008).

	animate singular	animate plural	inanimate singular	inanimate plural
agglutinating	flur-mu-pi	flur-mu-lo	flur-ka-pi	flur-ka-lo
fusional	flur-mu	flur-ka	flur-pi	flur-lo

**Table 1.** Illustration of the mapping of grammatical meanings onto suffixes in both conditions. *flur-* is the noun stem in each case, hyphens are provided for illustration only and did not appear in the stimuli.



**Figure 1:** Example trials in the agglutinating condition (left) and in the fusional condition (right) in Experiment 1, showing an inanimate plural trial and a singular animate trial, respectively.

## **Procedure**

Participants were instructed that they would be learning a part of a new language. The experiment was set up as an adaptive tracking paradigm where participants were not given any a-priori information about the language; the adaptive tracking paradigm provides trial-by-trial data, which tracks how learning unfolds over time. On each trial (**Figure 1**), participants saw an image and were given a choice of four words that could describe it. The four choices always consisted of the same stem, and differed only in which of the four possible affixes were attached to that stem (two affixes in the agglutinating, one affix in the fusional condition). The four choices referred to animate+singular, animate+plural, inanimate+singular and inanimate+plural. On each trial, only one choice was correct. Participants were instructed to click on the word that they thought correctly described the picture. Immediate feedback was given in every trial: if the participant's response was correct, the word they selected was highlighted in green; if it was incorrect their choice was highlighted in red and the correct word was highlighted in green; in either case the audio for the correct word was played aloud (audio was created using the Mac speech synthesizer with voice Alex). Participants' score was tracked throughout the experiment (1 point for each correct answer). The study consisted of 96 trials each of which displayed a unique picture. Therefore, no image or stem was ever repeated. This meant that participants were not required to learn the mappings between images and stems; their task was to learn to select the appropriate inflected form given the NUMBER and NOUN CLASS of the image presented. The combinations of grammatical meanings were balanced across the 96 trials so that each combination of meanings occurred as the correct choice 24 times in total.

At the end of the experiment, participants completed a short questionnaire asking for demographic information, language experience, and their experience during the experiment. The experiment lasted for approximately 25 minutes.

## **2.2 Results**

The experiment was designed to compare performance across conditions over time. Due to the nature of the adaptive tracking paradigm, participants were necessarily guessing at the beginning. We therefore expect performance to be similar across conditions early on, but to potentially diverge over the course of the experiment due to different rates of learning in the different conditions. Recall that above we predicted

that agglutinating languages should be easier to learn than fusional languages. This prediction should manifest as a steeper learning curve in this experiment; accuracy should increase earlier and faster in the agglutinating condition. **Figure 2** shows mean accuracy across conditions by trial. As expected, participants started at around chance level (25% correct, i.e. random selection among 4 options) and generally improved over trials. On average participants achieved high accuracy towards the end of the experiment in both conditions. However, contrary to our prediction, performance appeared to improve at a similar rate in both conditions. Mean accuracy across all trials for the agglutinating condition was 0.64 (SD=0.25), and the same for the fusional condition (0.64, SD=0.25). Across the final 10 trials, mean accuracy for the agglutinating condition was 0.75 (SD=0.30), for the fusional condition 0.73 (SD=0.31). To test whether the rate of improvement differed between conditions we fit a logistic polynomial mixed-effects regression model. The learning curves were modelled using second-order orthogonal polynomials, predicting correct answer by condition, trial, and their interaction. Since the polynomials were orthogonal, the model intercept refers to the overall average performance.<sup>6</sup> Trial was replaced by a linear and a quadratic time term; the time terms were added sequentially and the fit of the resulting two models (linear and polynomial) was compared using a Likelihood Ratio Test.

The inclusion of the quadratic time term significantly improved model fit ( $X^2=99.35$ ,  $p<0.001$ ). The polynomial model revealed a significant effect of the first time term ( $b=9.33$ ,  $SE=0.75$ ,  $p<0.001$ ) and of the second time term ( $b=-2.43$ ,  $SE=0.38$ ,  $p<0.001$ ), indicating that participants improved their accuracy over the course of the experiment and that their performance flattened out over time. However, there was no significant effect of condition ( $b=-0.02$ ,  $SE=0.16$ ,  $p=0.910$ ; NB this indicates the effect of condition when the two time terms are 0, i.e., at the midpoint of the experiment).

---

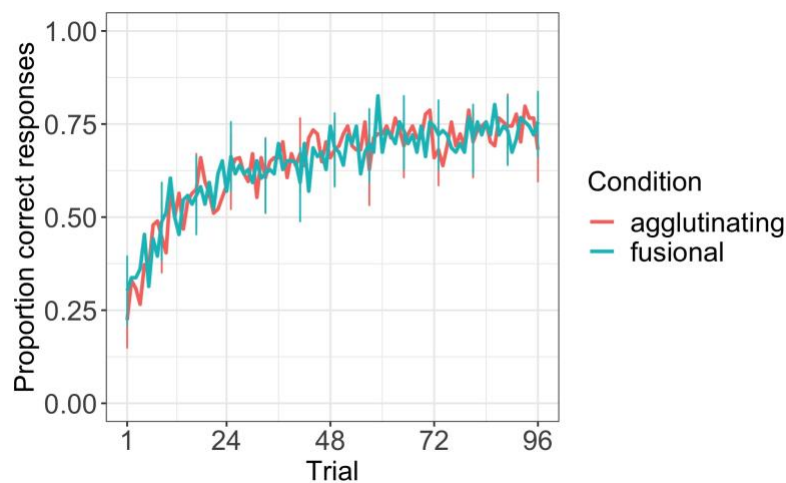
<sup>6</sup> All models were run using the package lme4 (Bates, 2015) in R (R Core Team, 2020). The specification of the polynomial models generally follows the procedure outlined in Mirman (2014: 44-60). We begin with a base model including all experimental predictors as well as a linear time term ot1. A subsequent model includes the quadratic time term ot2:

Model 1: `answer_correct ~ condition * ot1 + (1 + ot1 | participant) + (1 | item)`

Model 2: `answer_correct ~ condition * (ot1 + ot2) + (1 + (ot1 + ot2) | participant) + (1 | item)`

Models are compared using a Likelihood Ratio Test to evaluate whether the more complex model is justified by the data. Unless otherwise noted, models included random by-participant and by-item (picture) intercepts, and by-participant slopes for the time terms. The predictor condition (agglutinating vs. fusional) was effects-coded with agglutinating compared to the grand average. All models used the optimizer 'bobyqa'.

Importantly, there was no significant interaction effect between condition and the first time term ( $b=-0.10$ ,  $SE=0.70$ ,  $p=0.890$ ) or the second time term ( $b=-0.27$ ,  $SE=0.33$ ,  $p=0.415$ ). These non-significant interactions indicate that (i) there was no difference in the learning rate between conditions, and (ii) the learning curves flattened out over time similarly in both conditions (model results are also shown in **Table 2** in the Appendix). Thus, there is no indication of a learnability advantage of agglutinating over fusional morphology.



**Figure 2:** Mean accuracy by trial number and condition in Experiment 1. Error bars indicate 95% bootstrapped confidence intervals. Participants start at around chance level (0.25) on trial 1 and improve over time at a similar rate in both conditions.

### 2.3 Discussion

To summarise, in Experiment 1 we tested the hypothesis that transparency leads to better learnability by comparing participants' ability to learn an agglutinating system with separative exponence to a fusional system with cumulative exponence. It has been claimed that this makes agglutinating morphology easier to learn than fusional morphology for L1 and L2 learners (Don, 2017; Dressler, 2012; Goldschneider & DeKeyser, 2001; Igartua, 2015). We predicted that in our experiment the advantage for agglutinating systems would manifest as a steeper learning curve in the agglutinating condition. However, our results failed to confirm this prediction; we found no evidence for an advantage of agglutinating morphology over fusional morphology in overall learning.

One possible explanation for this apparent lack of advantage in our results is the difference in number of morphemes per word across conditions. As mentioned above, words in agglutinating languages tend to feature more morphemes per word, and this difference was present in our stimuli as well. In order to profit from the transparency of the agglutinating system used in our study, a learner must segment the suffixes from the stem. In the fusional condition, this means a single boundary must be identified between the stem and the single suffix. By contrast, the longer agglutinating words required segmentation of both suffixes. Because we fix morpheme length to one CV syllable across conditions, the words in the agglutinating condition are longer (again, a typical feature of agglutinating systems), meaning there are also more places where potential morpheme boundaries could fall. These factors may have led to more undersegmentation or mis-segmentation errors: for example, learners might initially treat the two suffixes as a single fusional morpheme, that is a single 4-letter morpheme which encodes both NUMBER and CLASS. If participants are undersegmenting in this way, no advantage of an agglutinating system would be expected. This issue may be exacerbated by the fact that learners in our experiment are adult second-language learners of this miniature artificial system. Even advanced L2 learners have been found to rely less on decomposition and more on whole-word lexical storage (Clahsen et al., 2010; Liang & Chen, 2014).

Analysis of the post-test questionnaire in Experiment 1 reveals that at least some participants reported failing to segment the stems and morphemes, or failed to segment the two agglutinating suffixes. While this pattern was observed in a minority of responses, it is nevertheless possible that segmentation issues indeed affected participants' speed of learning. Even if participants eventually segmented correctly, the effort of doing so may have masked any transparency advantage for the agglutinating system. In Experiment 2 we therefore test whether an advantage for the agglutinating system would be revealed in the presence of a visual cue to aid segmentation. In Experiment 3 we return to the issue of word length.

### **3. Experiment 2**

Experiment 2 used the same general design as in Experiment 1, but gave learners an explicit visual cue to segmentation. In particular, stems were presented in black, while suffixes were distinctly coloured: in the fusional condition the single fusional suffix was coloured, and in the agglutinating condition the two suffixes were each differently

coloured. If segmentation of the two suffixes presented a particular challenge in the agglutinating condition in the previous experiment, this manipulation should reduce or eliminate it, potentially allowing participants to exploit the transparency benefit of the agglutinating system.

### 3.1 Methods

#### Participants

100 participants were recruited online via Amazon Mechanical Turk, all self-reported to be native speakers of English, at least 18 years old, and possessed a Mechanical Turk qualification indicating they were based in the USA. They were each paid \$4 for their time. Participants were randomly assigned to one of the two conditions (49 to the agglutinating and 51 to the fusional condition).<sup>7</sup> We used the Mechanical Turk qualification and ID systems to ensure that participants were only able to participate in the experiment once; participants who participated in Experiment 1 were prevented from participating in Experiment 2 by the same means.

#### Materials

The language was identical to Experiment 1, however, a visual cue to the segmentation of stems and suffixes was provided as shown in **Figure 3**. As in the previous experiment, on each trial four options appeared, each with a different suffix or pair of suffixes. Stems were presented in black, and individual suffixes were highlighted with colour. This meant a single coloured cumulative morpheme in the fusional condition (either blue or orange, randomised across participants), and two differently coloured separative morphemes in the agglutinating condition (one blue, one orange, mapping randomised across participants).

---

<sup>7</sup> Five additional participants were excluded from analysis because they self-reported as bilingual. An additional 18 participants were excluded for having reported taking notes. The remaining data from 100 participants was used for the analysis. Of these, 20 participants indicated having at least basic skills in at least one other language apart from English.



**Figure 3:** Example trials in Experiment 2. The lefthand picture shows an inanimate, plural object in the agglutinating condition, with CLASS and NUMBER morphemes highlighted in two different colours. The righthand picture shows a singular inanimate object in the fusional condition, with fusional morphemes highlighted in one colour.

## Procedure

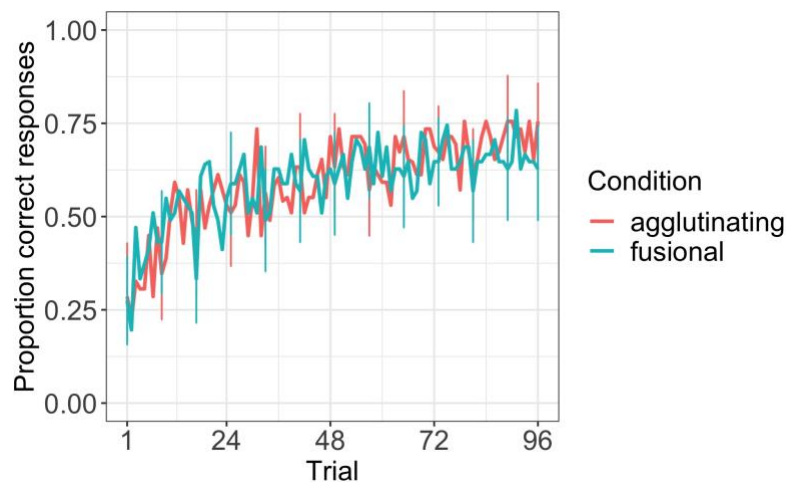
The procedure was identical to Experiment 1.

## 3.2 Results

**Figure 4** shows mean accuracy by trial across conditions. As in Experiment 1, participants started at around chance level (25%), improved over trials, and reached a high level of accuracy towards the end of the experiment in both conditions. As in the previous experiment, performance appears to improve at a very similar rate across conditions. Mean accuracy across all trials for the agglutinating condition was 0.59 (SD=0.25) and the same for the fusional condition (0.59, SD=0.23). Across the final 10 trials, mean accuracy for the agglutinating condition was 0.72 (SD=0.29), for the fusional condition 0.66 (SD=0.30).

As for Experiment 1, we fit a logistic polynomial mixed-effects regression model, using second-order orthogonal polynomials, predicting correct answer by condition, trial, and their interaction. The inclusion of the quadratic time term again significantly improved model fit ( $X^2=37.19$ ,  $p<0.001$ ). The polynomial model revealed a significant effect of the first time term ( $b=6.61$ ,  $SE=0.77$ ,  $p<0.001$ ) and of the second time term ( $b=-1.73$ ,  $SE=0.38$ ,  $p<0.001$ ), indicating that (i) accuracy improved over the course of the experiment and (ii) performance flattened out over time. There was no significant effect of condition ( $b=-0.03$ ,  $SE=0.18$ ,  $p=0.859$ ). Importantly, neither the interaction between condition and the first time term ( $b=0.52$ ,  $SE=0.73$ ,  $p=0.476$ ) nor the interaction between condition and the second time term ( $b=0.18$ ,  $SE=0.34$ ,

$p=0.597$ ) were significant. This indicates similar learning trajectories in both conditions (model results are also shown in **Table 3** in the Appendix). The results of Experiment 2 are therefore very similar to those of Experiment 1, with again no indication of a learnability advantage of agglutinating over fusional morphology.



**Figure 4:** Mean accuracy by trial number and condition in Experiment 2. Error bars indicate 95% bootstrapped confidence intervals. Participants start at around chance level (0.25) on trial 1 and improve over time at a similar rate in both conditions.

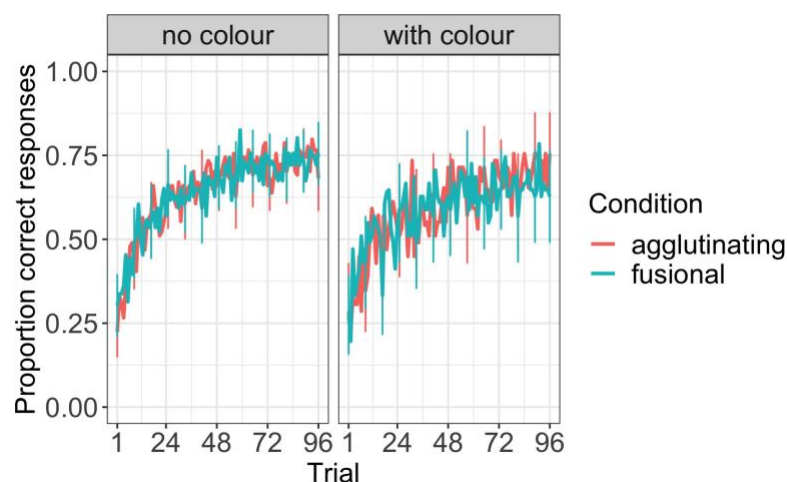
### 3.3. Combined analysis of Experiments 1 and 2

To assess whether colour in fact helped learning in either condition, we pooled the data from Experiments 1 and 2 together and repeated our analysis of overall accuracy. The dataset contained data from 280 participants (143 in the agglutinating and 137 in the fusional condition). We report the results in brief, full models are provided in Appendix **Table 4**.

**Figure 5** shows mean accuracy by trial across conditions in the combined dataset. We fit a logistic polynomial mixed-effects regression model, using second-order orthogonal polynomials, predicting correct answer by condition, trial, colour-coding, and their respective interactions.<sup>8</sup> The inclusion of the quadratic time term again significantly improved model fit ( $X^2 = 135.44$ ,  $p < 0.001$ ). The polynomial model revealed a marginal negative effect of colour-coding ( $b = -0.24$ ,  $SE = 0.12$ ,  $p = 0.051$ ) and a significant negative interaction between colour-coding and the first time term ( $b = -$

<sup>8</sup> The categorical predictor of colour-coding (two levels, not coloured [Experiment 1] vs. coloured [Experiment 2]) was effects-coded for better interpretability in this and the subsequent models. The levels agglutinating and coloured were thereby compared to the grand average in the models.

1.10, SE=0.53, p=0.039). Neither of these effects are in the predicted direction, indicating that if anything, adding colour cues to segmentation decreased performance in both conditions. There was again no effect of condition and no significant interaction between condition and any other predictor; crucially, there were no significant interactions involving colour-coding and condition which would indicate that colour-coding differentially affected performance in the agglutinating condition.



**Figure 5:** Mean accuracy by trial number, condition and colour-coding for the pooled data from Experiments 1 and 2. Error bars indicate 95% bootstrapped confidence intervals.

### 3.4 Discussion

In Experiment 2, we added a visual cue to segmentation in order to investigate the possibility that difficulty segmenting the two suffixes in the agglutinating condition might have masked any learning advantage for that type of system. However, the findings from Experiment 2 replicate those of Experiment 1: there was no evidence of a learnability advantage of agglutinating over fusional morphology. Further, analysis of the combined data suggests that colour was not particularly helpful to learning in this task, and did not differentially benefit participants in the agglutinating condition, which would have been predicted if the agglutinating condition posed particular challenges for segmentation which were alleviated by the colour cue. It is possible that participants found the colour cue distracting as an additional characteristic demanding attention. The enforcement of morpheme boundaries may have also made it more difficult to learn compared to focusing only on a subpart of the morphemes such as vowels. This strategy was reported by a few participants in Experiments 1 and 2. The

lack of an overall difference between conditions suggests that segmentation was generally not an issue for learners, and indeed that segmentation is not what prevents us from seeing better learning of agglutinating morphemes in the task.

Recall that our motivation for exploring the possible role of segmentation was a difference in the number of morphemes per word. As noted in passing earlier, this difference in the number of morphemes per word also leads to a difference in word length between the two conditions: our stimuli reflect a typological tendency for words with agglutinating morphemes to be longer than words with fusional morphemes. In particular, the agglutinating words were longer than the fusional words by one 2-letter morpheme. It is therefore possible that the additional length of the agglutinating words alone poses a cost to learning the agglutinating system. In Experiment 3, we tested this by matching the word length in the two conditions.

#### **4. Experiment 3**

Experiment 3 uses the same design as in Experiments 1 and 2, but we ran only the fusional condition, altering the fusional morphemes to make them 4 letters long. This creates an additional condition, matching the word length from the agglutinating conditions in the two previous experiments. We also manipulated colour highlighting, allowing for the possibility that word length and segmentation difficulty interact with one another.

##### **4.1 Methods**

###### **Participants**

200 participants were recruited online via Amazon Mechanical Turk, all self-reported to be native speakers of English, at least 18 years old, with a Mechanical Turk qualification indicating they were based in the USA. They were each paid \$4 for their time. All participants were assigned to the fusional condition.<sup>9</sup> We used the Mechanical Turk qualification and ID systems to ensure that participants were only able to

---

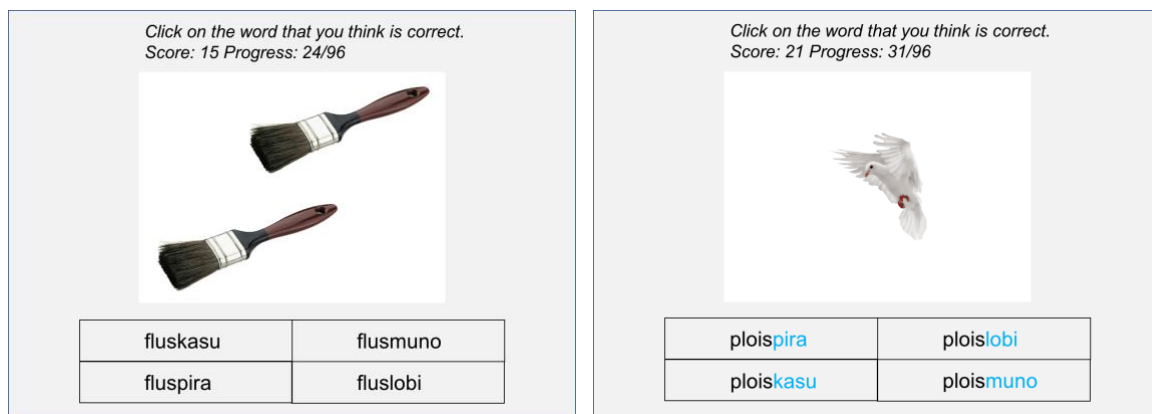
<sup>9</sup> Ten additional participants were excluded from analysis because they self-reported as bilingual. An additional 24 participants were excluded because they reported having taken notes during the experiment. The remaining data from 200 participants was used for the analysis. Of these, 39 participants indicated having at least basic skills in at least one other language apart from English.

participate in the experiment once, and that participants who participated in Experiments 1-2 were prevented from participating in Experiment 3.

## Materials

The language was identical to the fusional language from Experiments 1 and 2, except that we changed the fusional morphemes to each be 4 letters long and bisyllabic, to match the length and syllable structure of the grammatical endings of the agglutinating words in Experiments 1-2: the suffixes were *-muno*, *-kasu*, *-pira*, and *-lobi*. In this way, the resulting words were of the same length as the words in the previous agglutinating conditions, but they were still of the fusional type and could not be further segmented.

**Figure 6** shows example trials from each condition.



**Figure 6:** Example trials in Experiment 3. The lefthand picture shows a plural inanimate trial in the no-colour condition. The righthand picture shows a singular animate trial in the colour condition.

## Procedure

The procedure was identical to that of the previous experiments.

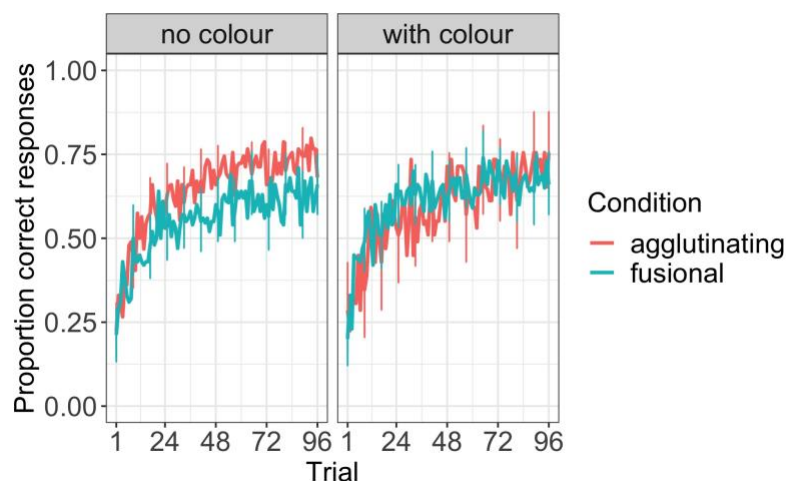
## 4.2 Results

**Figure 7** shows mean accuracy across conditions by trial. We compare the results from this experiment (200 participants) to those from the agglutinating conditions in Experiment 1 and 2 combined (143 participants).<sup>10</sup> Overall, as in the previous

<sup>10</sup> The data reported under Experiment 3 stem from altogether 4 experiments. We initially ran 100 participants in the length-matched fusional condition without (N=50) and with (N=50) colour highlighting, and saw the interaction between colour highlighting and morphological type reported in the analysis here. We then ran another 100 participants, replicating this result. We report the analysis of the pooled data for the purpose of brevity.

experiments, participants started at around chance level (0.25) and improved over trials. Mean accuracy across all trials in the agglutinating condition was 0.63 (SD=0.25), in the fusional condition 0.58 (SD=0.24). Across the last 10 trials, mean accuracy in the agglutinating condition was 0.74 (SD=0.30), in the fusional condition 0.65 (SD=0.31).

We again fit a logistic polynomial mixed-effects regression model using second-order orthogonal polynomials, predicting correct answer by condition, trial, colour-coding and their respective interactions. Random by-item intercepts were removed from both models due to convergence failure. The inclusion of the quadratic time term again significantly improved model fit ( $X^2 = 211.89$ ,  $p < 0.001$ ). In the resulting model, the effect of condition was not significant ( $b = 0.13$ ,  $SE = 0.10$ ,  $p = 0.227$ ); however, there was a significant interaction between condition and the first time term ( $b = 0.96$ ,  $SE = 0.42$ ,  $p = 0.022$ ) indicating that participants in the agglutinating condition improved more rapidly over time. **Figure 7** suggests this trend is mainly driven by the results from the agglutinating non-colour condition. There was no main effect of colour-coding, nor any significant interactions between colour and the time terms (see **Table 5** in the Appendix for the full results), indicating that the learning rates increased and flattened out similarly over time, independent of the presence of colour cues. However, the interaction between colour-coding and condition was significant ( $b = -0.23$ ,  $SE = 0.10$ ,  $p = 0.027$ ). This indicates that, at the midpoint of the experiment, participants in the agglutinating colour condition had a lower proportion of correct answers than in the agglutinating no-colour condition (see **Figure 7**). The interaction between condition, colour-coding and the second time term was also significant ( $b = 0.47$ ,  $SE = 0.23$ ,  $p = 0.043$ ), indicating that the learning curve for participants in the agglutinating colour condition flattened out less towards the end of the experiment compared to the average.



**Figure 7:** Mean accuracy by trial number, condition and colour-coding for the combined data from all conditions with 4-letter morphemes (agglutinating conditions from Experiments 1 and 2 and fusional conditions from Experiment 3). Error bars indicate 95% bootstrapped confidence intervals. Participants start at around chance level (0.25) on trial 1 and improve over time.

### 4.3 Interim discussion

In this experiment, we altered the length of the fusional suffixes in order to match the word length of the agglutinating conditions in Experiments 1 and 2. We then ran a combined analysis of the effects of condition and colour-coding. Notably, we found that when the length of the words (in this case their suffixes) is matched across conditions, the agglutinating system is indeed learned better over the course of the experiment. At first glance, this suggests that the reason why we did not see a learning advantage of the agglutinating condition in Experiments 1 and 2 may lie in the longer length of the inflected words in the agglutinating condition. As we noted before, this is an inherent feature of agglutinating systems, due to the recombination of individual morphemes to form words.

Crucially, however, a closer look at the role of colour-coding suggests that what we are seeing here is not a transparency advantage of agglutinating systems obscured by word length, but rather a difficulty in learning a fusional system with long morphemes. In particular, our results suggest that performance across conditions is modulated by colour-coding. When colour is present, we see no advantage of the agglutinating system. When colour is not present, there is a difference between the two conditions which is likely driven by segmentation problems in the length-matched fusional condition (see left-hand side of **Figure 7**). In particular, if participants in the

fusional condition initially attempt to segment the longer suffix at the syllable break within the one single morpheme (such as *mu-no*), they may be misled. Note that this strategy is plausible in the initial trials since the syllable structure matches that of the agglutinating system, where segmentation has been shown to happen at the syllable boundary even without additional colour cues. Thus, segmentation at the syllable break is in line with morpheme breaks in the agglutinating condition, but not in the fusional length-matched condition. This would explain why additional colour cues were not needed in Experiment 2, but improve performance in the fusional condition in Experiment 3. If this is indeed the case, then segmentation does play a role in how these systems are learned. In Experiment 2, we tested the hypothesis that *undersegmentation* in the agglutinating condition was obscuring any transparency advantage. But in fact what we see here suggests that *oversegmenting* longer fusional morphemes may be more problematic. Overall, these results also suggest that word length in and of itself is not detrimental to learning the agglutinating system.

Before we discuss what these results might tell us about the role of transparency in learning inflectional systems more generally, we first explore the types of errors participants made in the three experiments. As mentioned in footnote 1, an exploratory analysis of error patterns we conducted after running a first version of Experiment 1 suggested that participants in the fusional and agglutinating conditions might differ in the errors they make. Below we explore whether this difference holds across experiments, and if so what it might tell us about how transparency shapes learning.

#### **4.4 Analysis of error patterns**

Recall that the response choices participants are given in every trial constitute the four possible combinations of grammatical meanings: animate+singular, animate+plural, inanimate+singular, inanimate+plural. One of these responses was correct, and thus there were always three possibilities for an incorrect response: a participant could select the stem+suffix combination with the correct CLASS value but the incorrect NUMBER value; they could select the option with the correct NUMBER value but the incorrect CLASS value; or they could select the option with both feature values incorrect. Note of course that in the agglutinating condition these two features are expressed transparently in separate morphemes, whereas in the fusional condition both grammatical meanings are expressed by a single morpheme. Nonetheless, it is

possible to categorise all responses by the grammatical meanings on which they match the target response in both conditions. Thus, responses can be classified as either correct, CLASS correct only, NUMBER correct only, or none correct. We do not have any theoretically motivated prediction as to which of these options might be most likely, nor how the conditions might differ in terms of errors. However, an exploratory analysis of the data from a first version of Experiment 1 suggested the possibility that CLASS correct only errors were more common in the fusional condition. To investigate whether this difference was in fact replicated across the rest of our data here we report an analysis of error patterns across conditions. We first report results from the combined data of Experiments 1 and 2, and then for the suffix-length matched data from Experiment 3.

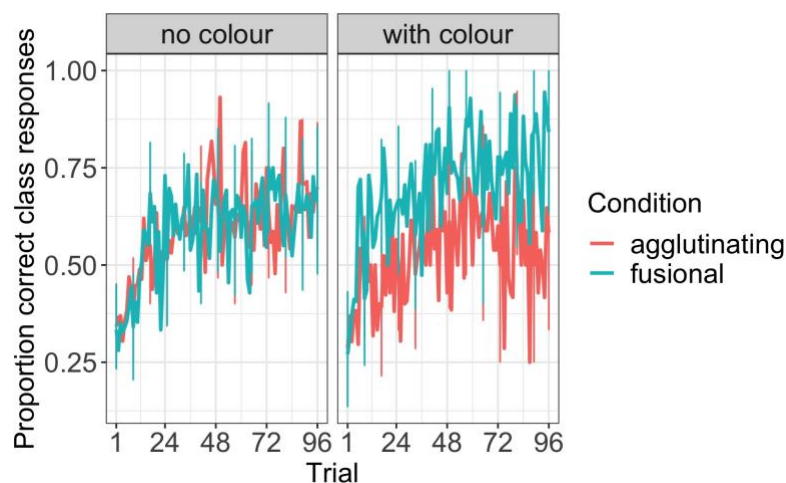
### **Combined analysis of data from Experiments 1 and 2**

Numerically, CLASS correct only responses were by far the most common, occurring about three times as often as each of the other two types of error (i.e. when prompted with a scene featuring e.g. an animate singular, error responses would tend to involve selecting the word form for animate plurals, rather than e.g. inanimate singulars). We see this consistently across all experiments and across the duration of the experiments. **Figure 8** isolates this effect, showing the proportion of CLASS correct errors among incorrect answers by trial in each condition for Experiments 1 (no-colour) and 2 (colour). We assessed the difference between conditions in CLASS correct errors using a logistic polynomial mixed-effects regression model fit to the subset of the data with incorrect answers. The dependent variable was CLASS correct. We followed the same procedure as outlined for the analysis of accuracy data above: we again used second-order orthogonal polynomials with fixed effects of condition, trial number (replaced by a linear and a quadratic time term), colour-coding and their respective interactions. Condition and colour-coding were again effects-coded, with agglutinating and colour being compared to the grand average.

The inclusion of the quadratic time term significantly improved model fit ( $X^2=112.96$ ,  $p<0.001$ ). The full model results are given in **Table 6** in the Appendix, here we focus on the effects of condition and colour. The polynomial model revealed a significant effect of condition ( $b=-0.33$ ,  $SE=0.10$ ,  $p=0.002$ ) and there was a significant interaction between condition and the first time term ( $b=-1.08$ ,  $SE=0.47$ ,  $p=0.021$ ). This indicates that CLASS correct only errors were found more in the fusional

condition than in the agglutinating condition and that across trials, CLASS correct only errors reduced significantly more rapidly in the agglutinating condition compared to the grand mean. There was no effect of colour ( $b=0.05$ ,  $SE=0.10$ ,  $p=0.653$ ), but a significant interaction between colour-coding and condition ( $b=-0.21$ ,  $SE=0.10$ ,  $p=0.045$ ), suggesting that the likelihood of CLASS correct only errors may be differently affected by colour in the two conditions: the agglutinating colour condition yields significantly fewer CLASS correct only errors compared to the grand average. We return to this effect further below.

Overall, these results suggest that participants in the two conditions diverged over time in their tendency to choose an answer in which only the CLASS morpheme was correct: the proportion of such errors declines faster in the agglutinating condition and the conditions differ more from each other in the presence of colour cues. This trend can be observed clearly in **Figure 8**, where the proportion of CLASS correct only errors reduces more for the agglutinating than the fusional condition when colour is present. While we did not observe an effect of condition with respect to overall accuracy in Experiments 1 and 2, the analysis of error patterns suggests that participants' learning strategies may differ between conditions. We will return to this (and the interaction with colour) below.

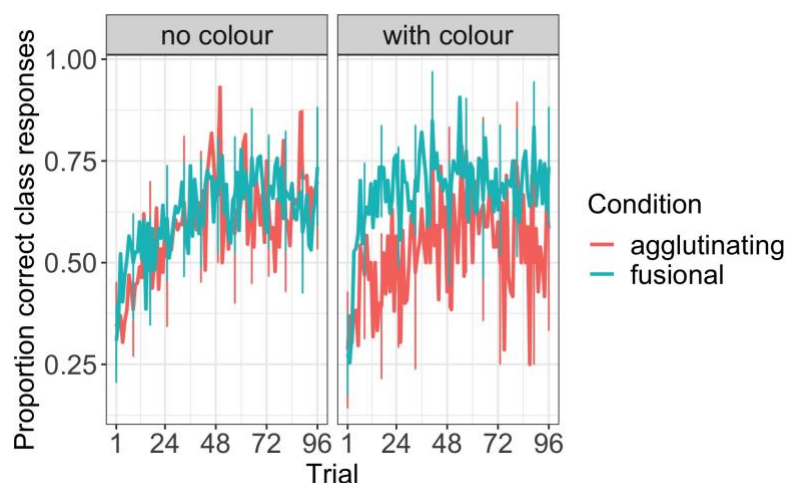


**Figure 8:** Mean proportion of incorrect responses in which only CLASS was correct, by trial number, condition and colour-coding for the pooled data from Experiments 1 and 2. Error bars indicate 95% bootstrapped confidence intervals. The figure displays only the subset of incorrect answers.

### Analysis of CLASS correct errors in suffix-length-matched conditions

We repeated our analysis of CLASS correct errors using the length-matched agglutinating and fusional conditions. **Figure 9** shows the proportion of incorrect responses in which CLASS was correct for the combined data from all conditions with suffixes of length 4 (agglutinating conditions from Experiments 1-2 and fusional conditions from Experiment 3). We again used second-order orthogonal polynomials predicting correct CLASS in the subset of the data with incorrect answers, including fixed effects of condition, trial, colour-coding and their interaction. The inclusion of the quadratic time term again significantly improved model fit ( $X^2 = 144.75$ ,  $p < 0.001$ ). The effect of condition at the midpoint of the experiment was significant ( $b = -0.35$ ,  $SE = 0.10$ ,  $p < 0.001$ ), indicating that, as above, participants in the agglutinating condition focused on CLASS less compared to the average. While **Figure 9** suggests this effect was largely driven by the colour condition (as was the case in the error analysis of Experiments 1-2), for Experiment 3 the interaction between condition and colour was not significant ( $b = -0.13$ ,  $SE = 0.1$ ,  $p = 0.17$ ). The interaction between colour-coding and the first time term was marginally significant ( $b = -0.72$ ,  $SE = 0.40$ ,  $p = 0.075$ ), indicating that CLASS correct only responses decreased slightly over time when colour cues were present (see **Table 7** in the Appendix for the full results).

Overall, our results for Experiment 3 suggest that participants learning the agglutinating system focus less on CLASS than those learning the fusional system, which is in line with the results of the error patterns from Experiments 1 and 2.



**Figure 9:** Mean proportion of incorrect responses in which CLASS was correct, by trial number, condition and colour-coding for the combined data from all conditions with 4-letter suffixes (agglutinating conditions from Experiments 1-2 and fusional conditions

from Experiment 3). Error bars indicate 95% bootstrapped confidence intervals. The figure displays only the subset of incorrect answers.

To summarise, our post-hoc analysis of error types suggests that participants in the fusional condition make more CLASS correct only errors. One possibility is that this pattern reflects the kind of bias for transparency that we predicted to affect overall learnability: participants in the fusional condition may have been searching for a single feature with four values, rather than two binary features. While NUMBER is unambiguously binary in our stimuli (one vs. two), the stimuli could in principle encode more fine-grained distinctions of CLASS. In the post-test questionnaire, some participants indeed reported such a strategy, for instance, distinguishing land vs. sea animals and household items vs. clothing, attempting to map each of these four CLASS values onto one morpheme, ignoring NUMBER altogether. While it is clearly imperfect, such a strategy would have often led participants to pick the correct CLASS morpheme, since, for instance, land and sea animals are both animate. It would also explain why the focus on CLASS is stronger in the fusional condition: assuming transparency of morphemes in the agglutinating condition would lead to the correct hypothesis that each morpheme carries the meaning of a single feature.

Adding colour cues in the agglutinating condition does not increase the focus on CLASS across trials, but this is not something we would necessarily expect either. The observed difficulty with CLASS seems to lie in the fusional condition and it is not clear how colour highlighting could help learners pick up on the number of features mapped onto each morpheme in the fusional condition. Likewise, it appears as if participants were already segmenting the two agglutinating morphemes even without colour (see overall analysis of Experiments 1 and 2), which should allow for a successful mapping of the features onto the morphemes even without an additional visual aid. However, as noted, there is a significant difference between conditions at the midpoint of the experiment as a function of colour: participants learning the fusional system with colour cues tend to focus on CLASS the most (**Figures 8 and 9**). A possible explanation for this is that colour reinforces the morpheme boundaries, which makes performance diverge across conditions: the reinforcement of a single morpheme seems to increase the bias for transparency in the fusional condition, and to support the correct mapping between meaning and form in the agglutinating condition.

## 5. General Discussion

The experiments we present in this paper were conducted to test the hypothesis that agglutinating morphemes should be easier to learn than fusional morphemes due to their relative transparency. Perhaps the defining feature of agglutinating morphemes is their one-to-one correspondence between meaning and form, called separative exponence (Plank, 1986; Plungian, 2008). In contrast, fusional morphemes typically display cumulative exponence, which is a many-to-one mapping between meaning and form. The relative transparency of agglutinating morphology has been argued to lead to better learnability in children and adults, both in L1 and L2 acquisition (e.g., Bittner et al., 2003; DeKeyser, 2005; Narasimhan & Gullberg, 2011). Instances of language change have also been ascribed to a preference for transparent structures (Anttila, 1989; Igartua, 2015). However, transparency is only one of the features that distinguishes agglutinating from fusional systems in natural languages. Agglutinating systems also typically exhibit more regularity (i.e., fewer allomorphs, syncretisms etc.), the compositional nature of agglutinating systems means they have the advantage of allowing for generalisation to new meanings, and individual morphemes repeat more often. Here we used an artificial language learning paradigm in order to isolate, as far as possible, the effect of transparency.

In Experiment 1, we set up a two-by-two paradigm, with morphemes expressing NUMBER (singular, plural) and CLASS (animate, inanimate). Conditions were matched on the number of morphemes to be learned but differed in whether each feature value was expressed by a single morpheme (agglutinating condition), or whether an individual morpheme expressed both feature values (fusional condition). We failed to find a learning advantage of agglutinating over fusional morphology. In Experiment 2, we added an explicit cue to morpheme boundaries using colour, since in principle the agglutinating condition involves a greater burden of segmentation (two morpheme boundaries rather than one). However, this failed to reveal any advantage of agglutinating morphology, suggesting that segmentation by itself did not pose a major difficulty for participants in this task. In Experiment 3, we increased the length of the morphemes in the fusional condition such that the total word length was matched across conditions, and we again tested participants with and without an explicit cue to morpheme boundaries. Comparing these two new fusional conditions with the agglutinating conditions in Experiments 1 and 2, we found that the agglutinating system was learned better than the fusional one, but only when no cue to morpheme

boundaries was provided. This effect appears to be driven by segmentation difficulty in the fusional condition rather than by any transparency advantage of the agglutinating system.

While we failed to find the predicted overall advantage for the agglutinating condition, participants' errors revealed a pattern which may nevertheless indicate a bias for transparency. In particular, in the fusional condition, participants' errors were more likely to involve choosing an option which expressed the correct CLASS value, but the incorrect NUMBER value. While NUMBER was unambiguously binary in our task (one vs. two), the stimuli could in principle be analysed as containing more fine-grained distinctions of CLASS. Indeed, post-experiment questionnaire responses indicated that many participants ignored NUMBER and instead attempted to find 4 sub-categories within CLASS (e.g., land animals, sea animals, household items, clothing), even though they sometimes noticed that their CLASS-based strategy was not fully successful.

We find it surprising that participants would ignore NUMBER in our task in favour of pursuing a 4-way distinction based on CLASS, since NUMBER is visually very salient. However, it is clear that animacy distinctions are very salient to humans (e.g., Blumenthal et al., 2018; Bonin et al., 2019; Calvillo & Jackson, 2014; Kriegeskorte et al., 2008; Strickland et al., 2017). On top of that, there is evidence that adult learners sometimes ignore additional informative cues when multiple cues are present but one is more salient (the so-called 'cue competition effect', e.g., Soto & Wasserman, 2010). Relatedly, adults show effects of selective attention, in that they have difficulty considering cues they had previously ignored (e.g., Blanco & Sloutsky, 2019). These findings suggest the possibility that once participants in our task identified animacy as relevant, they may have found it difficult to subsequently consider other criteria they had previously ignored, in our case NUMBER. Critically, that this occurred more in the fusional condition suggests that a transparency bias is also at work. In the fusional condition, imposing a 4-way categorisation within CLASS onto the four fusional morphemes while ignoring NUMBER creates a transparent system from a non-transparent one. In the agglutinating condition, the morphemes are already transparent, and therefore imposing this kind of strategy is not motivated by a transparency bias. Our analysis shows that this strategy was particularly common for participants in the fusional condition in the presence of an explicit cue to morpheme boundaries (**Figure 8**). In the fusional condition, this cue reinforces that there is only a single morpheme, potentially strengthening the motivation for using a 4-way CLASS

distinction. That said, if this pattern of results is indeed explained in part by a transparency bias, this bias is strong enough to make participants ignore NUMBER, but at the same time too weak to lead to any overall difference in learning across conditions.

In the remainder of this section, we explore what we might conclude from this failure to find an effect of transparency as predicted, and what it tells us about the acquisition of agglutinating versus fusional morphemes. We see three general possibilities, which we will discuss in turn. First, perhaps our experiment was not able to tap into the types of pressures or biases that really face first and second language learners. Second, perhaps agglutinating systems present additional difficulties that are not necessarily overcome by a transparency advantage, at least not in our task. Third, perhaps the benefits of agglutinating morphology only appear once the system is sufficiently large.

The first possibility, and perhaps the least interesting one, is simply that artificial language learning experiments are non-naturalistic in some critical way which obscures the true transparency advantage of agglutinating systems. Artificial language learning experiments have a rich history in developmental psychology and linguistics, and have been successfully used in exactly the kinds of circumstances we use them here: where we are interested in comparing the learnability of different systems to explore the basis of some typological pattern of interest (e.g., for reviews of this literature see Culbertson, 2012; Culbertson & Schuler, 2019; Levshina, 2019; Moreton & Pater, 2012a,b). Especially since transparency typically co-occurs with additional confounding factors in natural languages, an artificial language learning paradigm is an ideal way of isolating the effect of transparency. Further, differences in the learnability of morphological systems have been argued to hold in both first and second language learners, therefore there is no reason to suspect that using adult participants has somehow obscured the bias of interest. That said, our participants were native speakers of English, a language without a rich inflectional system, and it could be that they are not particularly good at learning inflections. This is something we will explore in future work. However, it is worth noting that on average participants are well above chance levels of accuracy by the end of our task. Reversely, the lack of a difference between our conditions is not due to participants being at ceiling. Importantly, our study was designed to allow us to look at the trajectory of learning, not just the endpoint, and average accuracy across all trials was around 60 percent.

We therefore set experimental design aside for the moment and consider two further possible explanations.

Above, we described several apparent advantages of agglutinating systems – including generalisability, frequency and regularity of morphemes across the system, and transparency of the mapping from meaning to form – which are typically confounded in natural languages. Our motivation here was to isolate as much as possible the effect of transparency. But one can imagine some otherwise more minor difficulties of agglutinating systems which might come into play when the other advantages are not all present. For example, while this does not seem to have been an issue in our task, longer words and more morpheme boundaries are nevertheless an in-principle cost of agglutinating systems. Similarly, learning and retrieval of the form-meaning mappings is, all things equal, more difficult for agglutinating systems. In our task, even once a participant has successfully categorised the stimuli and learned which morphemes map to each feature value, they will need to retrieve both correct morphemes on any given trial. In that sense, even though the number of morphemes is the same across conditions, there is more material to recall on each trial in the agglutinating than in the fusional condition. In principle, the correct order of morphemes also needs to be acquired, though we do not require participants in the task to produce morpheme sequences, so this is less likely to be at play (and seems not to cause much trouble during acquisition, e.g., Leonard, 2014; Slobin, 1973). While in principle the cost of learning and retrieving the mappings between meanings and morphemes might obscure the transparency advantage of the agglutinating system, we believe there is a more likely explanation at play.

In particular, we may not see a learnability difference between agglutinating and fusional systems in our experiments due to the small size of the paradigm (or meaning space) we tested. There are two possible reasons for this. First, it may be that the transparency advantage exists, but is weak, and therefore would only be revealed if the paradigms to be learned were larger. In other words, learning a handful of fusional morphemes, even if they are not transparent, is simply not that hard. The second possibility is that the advantage of agglutinating systems lies not (or at least not mainly) in transparency, but instead is due to the benefits of compositionality and related factors.

There is indeed some typological evidence suggesting that the benefits of agglutination are confined to larger paradigms. Plank (1986) proposes a link between

the type of exponence and the number of feature values in inflectional categories, namely that systems with cumulative exponents tend to have fewer feature values than those with separative exponents. Crucially, however, he adds that “It does not follow [from his suggested link] that any type of exponence is less suitable than any other to express categories *of relatively few terms*” (Plank, 1986, pp.33-34, our italics). Indeed, in a typological study of case marking in 47 languages, Plank (1986) observes that the number of languages that either express case separatively or cumulatively is similar for languages with few case terms, but in languages which encode many case distinctions, separative exponence prevails; no cumulative language in the sample has more than 10 case terms, and the languages with more than 10 case terms (e.g. Basque, Evenki, Hungarian, Estonian, Finnish) are all separative. This is supported by Laaha & Gillis (2007), who report that in a sample of 9 languages, those classified as strongly fusional express fewer features on verbs than those classified as strongly agglutinating. These typological findings suggest the possibility that when the category system is small, as in our experiments, both types of exponence are similarly common because they are similarly learnable.

There is also evidence from computational modelling supporting the idea that differences in learnability across small paradigms might be negligible. Recall that compositional systems allow faster generalisation: once you have encountered (and learned) the morpheme for one feature value, you can generalise that to any other item with that feature. However, the advantage of a compositional system over a holistic one likely depends on the size of the paradigm. For example, Brighton (2002) calculates that the generalisation advantage of compositionality over holistic systems only occurs when the meaning space reaches a certain size (level of complexity in his terms). A compositional system with only two binary features only has a slight advantage over a holistic system of the same size in his model; indeed, this motivated our use of 2x2 paradigms, since we wanted to isolate the putative advantage of transparency in agglutinating systems. It is worth noting that the model used by Brighton (2002) involves an ideal learner that only requires exposure to a particular expression once to learn. Further, the learner only sees a subset of possible meanings in the meaning space, and the advantage of one system over another depends purely on how successfully the learner can use it to express new meanings (i.e., to generalise). These results do not necessarily map on to our experiments straightforwardly. Our participants do not (necessarily) successfully learn the

morphemes based on a single exposure, and they were exposed to the entire morphological system within relatively few trials. However, Brighton's model captures the intuition that participants in the agglutinating condition will cycle through all four of the semantic feature values in a relatively short time, multiple times throughout the experiment. By contrast, each cycle through the feature set will take longer in the fusional condition. Considered this way, although compositionality may play some small role in our experiment, the results of Brighton (2002) suggest that the real advantages of an agglutinating system only come into play when the system is sufficiently large.

Setting up an experiment with a larger paradigm (or meaning space) and incorporating generalisation into the task could allow us to explore whether and how paradigm size interacts with the learnability of agglutinating and fusional systems. However, larger paradigms will necessarily reintroduce confounds which we sought to avoid here. For example, even without requiring learners to generalise to new meanings, a larger paradigm will mean that the higher frequency of agglutinating morphemes relative to fusional morphemes may have an impact on learning. Further, in larger paradigms, it is not possible to hold the number of morphemes constant across the two types of systems. For example, a three-by-three system (which shows a clear advantage for compositional systems in Brighton, 2002) requires 9 agglutinating morphemes but 27 fusional morphemes, meaning that any learnability advantage for agglutinating systems may derive from the smaller inventory of morphemes they require, rather than their transparency. Finally, fusional morphology often displays more variation than agglutinating morphology, for instance stem and affix variation and inflectional classes. This contrasts with the regularity of agglutinating systems and might constitute a forth reason why agglutinating systems should be easier to learn.

To summarise, using a small, two-by-two paradigm has obvious benefits, and yet may be too small to reveal any advantage, either related to transparency or compositionality. We would argue that the failure to find a clear effect of transparency under these conditions suggests that the impact of transparency in morphological learning may be overstated. If transparency in isolation had a significant effect on the learnability of morphological systems, then we should see it in our task. In this sense, our findings support the idea that the learnability of morphological systems stems from

the combined effects of multiple factors, some of which may not always be at play (Andersen, 1978; Goldschneider & DeKeyser, 2001; Penke, 2012).

## **6. Conclusion**

A common claim over the past decades has been that agglutinating morphological systems are easier to learn than fusional systems (during both first and second language acquisition) because they are transparent. Since the transparency of exponents in natural languages is typically correlated with other factors such as compositionality, generalisability and frequency of the forms in question, we devised an artificial language learning paradigm that allows us to test the effects of transparency on learning in isolation. However, across three separate experiments we find that the transparency of morphological systems alone does not lead to significant differences in learnability. We also find that our results are not due to larger segmentation effort or longer words in agglutinating systems, which could potentially pose disadvantages to a learner. The most likely explanation for our results seems to be that small transparent agglutinating systems do not provide much benefit over small fusional systems. We expect the advantage to appear when the systems and their differences become larger, in which case transparency correlates with other factors, as has been found in several previous studies. Therefore, the impact of transparency on learnability may be overstated and instead, the difference in learnability of inflectional elements may be best described by considering the interplay of transparency with factors like generalisability and frequency.

## References:

- Aksu-Koç, A., & Slobin, D. (1985). The acquisition of Turkish. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 1, pp. 839–878). Hillsdale, NJ: Lawrence Erlbaum.
- Andersen, R. W. (1978). AN IMPLICATIONAL MODEL FOR SECOND LANGUAGE RESEARCH. *Language Learning*, 28(2), 221–282.  
<https://doi.org/10.1111/j.1467-1770.1978.tb00134.x>
- Anttila, R. (1989). *Historical and Comparative Linguistics*. John Benjamins Publishing Company.
- Bates D., Mächler M., Bolker B., & Walker S. (2015). “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01
- Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution*, 2(2), 160–176. <https://doi.org/10.1093/jole/lzx001>
- Bickel, B., & Nichols, J. (2013). Exponence of Selected Inflectional Formatives. In: Dryer, Matthew S. & Haspelmath, Martin (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/21>, Accessed on 2021-04-29.)
- Bittner, D., Dressler, W., & Kilani-Schoch, M. (2003). *Development of verb inflection in first language acquisition. A cross-linguistic perspective*. deGruyter.
- Blanco, N. J., & Sloutsky, V. M. (2019). Adaptive flexibility in category learning? Young children exhibit smaller costs of selective attention than adults. *Developmental Psychology*, 55(10), 2060–2076.  
<https://doi.org/10.1037/dev0000777>
- Blumenthal, A., Stojanoski, B., Martin, C. B., Cusack, R., & Köhler, S. (2018). Animacy and real-world size shape object representations in the human medial temporal lobes. *Human Brain Mapping*, 39(9), 3779–3792.  
<https://doi.org/10.1002/hbm.24212>
- Bonin, P., Gelin, M., Dioux, V., & Méot, A. (2019). “It is alive!” Evidence for animacy effects in semantic categorization and lexical decision. *Applied Psycholinguistics*, 40(4), 965–985.  
<https://doi.org/10.1017/S0142716419000092>
- Brighton, H. (2002). Compositional Syntax From Cultural Transmission. *Artificial Life*, 8(1), 25–54. <https://doi.org/10.1162/106454602753694756>
- Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, 2(3), 177–226.  
<https://doi.org/10.1016/j.plrev.2005.06.001>
- Brown, R. (1976). *First Language: The Early Stages*. Harvard University Press.
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. J. Benjamins.
- Calvillo, D. P., & Jackson, R. E. (2014). Animacy, perceptual load, and inattention blindness. *Psychonomic Bulletin & Review*, 21(3), 670–675.  
<https://doi.org/10.3758/s13423-013-0543-8>
- Chaabouni, R., Kharitonov, E., Bouchacourt, D., Dupoux, E., & Baroni, M. (2020). Compositionality and Generalization in Emergent Languages. *ArXiv:2004.09124 [Cs]*. <http://arxiv.org/abs/2004.09124>

- Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological Structure in Native and Nonnative Language Processing. *Language Learning*, 60(1), 21–43. <https://doi.org/10.1111/j.1467-9922.2009.00550.x>
- Culbertson, J. (2012). Typological Universals as Reflections of Biased Learning: Evidence from Artificial Language Learning: Typological Universals as Reflections of Biased Learning. *Language and Linguistics Compass*, 6(5), 310–329. <https://doi.org/10.1002/lnc3.338>
- Culbertson, J., & Schuler, K. (2019). Artificial Language Learning in Children. *Annual Review of Linguistics*, 5(1), 353–373. <https://doi.org/10.1146/annurev-linguistics-011718-012329>
- DeKeyser, R. M. (2005). What Makes Learning Second-Language Grammar Difficult? A Review of Issues: What Makes Learning L2 Grammar Difficult? *Language Learning*, 55(S1), 1–25. <https://doi.org/10.1111/j.0023-8333.2005.00294.x>
- de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, 2(3), 267–278. <https://doi.org/10.1007/BF01067106>
- Don, J. (2017). What causes languages to be transparent? *Language Sciences*, 60, 133–143. <https://doi.org/10.1016/j.langsci.2017.01.004>
- Dressler, W. U. (2005). Morphological Typology and First Language Acquisition: Some Mutual Challenges. *On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4)*. URL: <http://morbo.lingue.unibo.it/mmm/>
- Dressler, W. U. (2012). On the acquisition of inflectional morphology: Introduction. *Morphology*, 22(1), 1–8. <https://doi.org/10.1007/s11525-011-9198-1>
- Ellis, N., & Collins, L. (2009). Input and Second Language Acquisition: The Roles of Frequency, Form, and Function Introduction to the Special Issue. *The Modern Language Journal*, 93(3), 329–335. <https://doi.org/10.1111/j.1540-4781.2009.00893.x>
- Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the “Natural Order of L2 Morpheme Acquisition” in English: A Meta-analysis of Multiple Determinants. *Language Learning*, 51(1), 50.
- Haspelmath, M. (2009). An Empirical Test of the Agglutination Hypothesis. In S. Scalise, E. Magni, & A. Bisetto (Eds.), *Universals of Language Today* (Vol. 76, pp. 13–29). Springer Netherlands. [https://doi.org/10.1007/978-1-4020-8825-4\\_2](https://doi.org/10.1007/978-1-4020-8825-4_2)
- Hengeveld, K., & Leufkens, S. (2018). Transparent and non-transparent languages. *Folia Linguistica*, 52(1), 139–175. <https://doi.org/10.1515/flin-2018-0003>
- Igartua, I. (2015). From cumulative to separative exponence in inflection: Reversing the morphological cycle. *Language*, 91(3), 676–722. <https://doi.org/10.1353/lan.2015.0032>
- Johnson, T., Gao, K., Smith, K., Rabagliati, H., & Culbertson, J. (2020). *Predictive structure or paradigm size? Investigating the effects of i-complexity and e-complexity on the learnability of morphological systems* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/qns6v>
- Kempe, V., & Brooks, P. J. (2008). Second Language Learning of Complex Inflectional Systems. *Language Learning*, 58(4), 703–746. <https://doi.org/10.1111/j.1467-9922.2008.00477.x>
- Kirby, S. (2000). Syntax Without Natural Selection: How Compositionality Emerges from Vocabulary in a Population of Learners. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *The Evolutionary Emergence of Language* (1<sup>st</sup>

- ed., pp. 303–323). Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511606441.019>
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686. <https://doi.org/10.1073/pnas.0707835105>
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102. <https://doi.org/10.1016/j.cognition.2015.03.016>
- Korecky-Kröll, K., & Dressler, W. U. (2009). The acquisition of number and case in Austrian German nouns. In U. Stephany & M. D. Voeikova (Eds.), *Development of Nominal Inflection in First Language Acquisition* (Vol. 30, pp. 265–302). Mouton de Gruyter.  
<https://doi.org/10.1515/9783110217117.265>
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–1141.
- Krifka M. (2001). Compositionality. In: Wilson RA, Keil F, editors. *The MIT encyclopaedia of the cognitive sciences*. Cambridge, MA: MIT Press.
- Laaha, S., & Gillis, S. (Eds.) (2007). *Typological perspectives on the acquisition of noun and verb morphology. Antwerp Papers in Linguistics.*
- Larsen-Freeman, D. (2010). Not So Fast: A Discussion of L2 Morpheme Processing and Acquisition. *Language Learning*, 60(1), 221–230.  
<https://doi.org/10.1111/j.1467-9922.2009.00556.x>
- Leonard, L. B. (2014). Specific Language Impairment Across Languages. *Child Development Perspectives*, 8(1), 1–5. <https://doi.org/10.1111/cdep.12053>
- Levshina, N. (2019). *Linguistic Frankenstein, or How to test universal constraints without real languages*. Zenodo. <https://doi.org/10.5281/ZENODO.2583820>
- Liang, L., & Chen, B. (2014). Processing morphologically complex words in second-language learners: The effect of proficiency. *Acta Psychologica*, 150, 69–79.  
<https://doi.org/10.1016/j.actpsy.2014.04.009>
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. CRC Press/Taylor & Francis Group.
- Moreton, E., & Pater, J. (2012a). Structure and Substance in Artificial-phonology Learning, Part I: Structure: Structure and Substance in Artificial-Phonology Learning, Part I. *Language and Linguistics Compass*, 6(11), 686–701.  
<https://doi.org/10.1002/lnc3.363>
- Moreton, E., & Pater, J. (2012b). Structure and Substance in Artificial-Phonology Learning, Part II: Substance: Structure and Substance in Artificial-Phonology Learning, Part II. *Language and Linguistics Compass*, 6(11), 702–718.  
<https://doi.org/10.1002/lnc3.366>
- Narasimhan, B., & Gullberg, M. (2011). The role of input frequency and semantic transparency in the acquisition of verb meaning: Evidence from placement verbs in Tamil and Dutch. *Journal of Child Language*, 38(3), 504–532.  
<https://doi.org/10.1017/S0305000910000164>
- Penke, M. (2012). *The Acquisition of Inflectional Morphology*.  
<https://doi.org/10.13140/RG.2.2.21194.06085>
- Pfänder, S., Behrens, H., Auer, P., Jacob, D., Kailuweit, R., Konieczny, L., Kortmann, B., Mair, C., & Strube, G. (2013). Erfahrung zählt. Frequenzeffekte

- in der Sprache – ein Werkstattbericht. *Zeitschrift für Literaturwissenschaft und Linguistik*, 43(1), 7–32. <https://doi.org/10.1007/BF03379870>
- Plank, F. (1986). PARADIGM SIZE, MORPHOLOGICAL TYPOLOGY, AND UNIVERSAL ECONOMY. *San Diego*, 20.
- Plungian, V. A. (2008). Agglutination and flection. In Haspelmath, M., Koenig, E., Oesterreicher, W. (Eds). *Language Typology and Universals*. (pp.669-678). DeGruyter. <https://doi.org/10.1515/9783110194036-012>
- R Core Team (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.) *Studies of Child Language Development*. Holt, Rinehart & Winston.
- Soto, F. A., Wasserman, E. A. (2010). Missing the forest for the trees: Object discrimination learning blocks categorization learning. *Psychological Science*, 21(10), 1510–1517.
- Strickland, B. (2017). Language reflects “core” cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive science*, 41(1):70–101.
- Vogt, P. (2005). On the Acquisition and Evolution of Compositional Languages: Sparse Input and the Productive Creativity of Children. *Adaptive Behavior*, 13(4), 325–346. <https://doi.org/10.1177/105971230501300403>
- Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., Gagarina, N., Hrzica, G., Ketrez, F. N., Kilani-Schoch, M., Korecky-Kröll, K., Kovac̆evic̆, M., Laalo, K., Palmovic̆, M., Pfeiler, B., Voeikova, M. D., & Dressler, W. U. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Language*, 31(4), 461–479. <https://doi.org/10.1177/0142723711409976>

## Appendix

### 1. Full model results – overall accuracy

	Estimate	Std. Error	z value	Pr(>   z   )
(Intercept)	1.43	0.16	8.90	< 0.001
condition1	-0.02	0.16	-0.11	0.910
ot1	9.33	0.75	12.37	< 0.001
ot2	-2.43	0.38	-6.45	< 0.001
condition1:ot1	-0.10	0.70	-0.14	0.890
condition1:ot2	-0.27	0.33	-0.81	0.415

**Table 2:** Results from Experiment 1, overall accuracy. The predictor of condition was effects coded with the agglutinating condition being compared to the grand average.

	Estimate	Std. Error	z value	Pr(>   z   )
(Intercept)	0.87	0.18	4.88	< 0.001
condition1	-0.03	0.18	-0.18	0.859
ot1	6.61	0.77	8.59	< 0.001
ot2	-1.73	0.38	-4.51	< 0.001
condition1:ot1	0.52	0.73	0.71	0.476
condition1:ot2	0.18	0.34	0.53	0.597

**Table 3:** Results from Experiment 2, overall accuracy. The predictor of condition was effects coded with the agglutinating condition being compared to the grand average.

	Estimate	Std. Error	z value	Pr(>   z  )
(Intercept)	1.15	0.13	9.21	< 0.001
condition1	-0.03	0.12	-0.23	0.819
ot1	7.97	0.56	14.21	< 0.001
ot2	-2.12	0.28	-7.53	< 0.001
colorcoding1	-0.24	0.12	-1.95	0.051
condition1:ot1	0.19	0.53	0.36	0.717
condition1:ot2	-0.03	0.25	-0.13	0.898
condition1:colorcoding1	-0.01	0.12	-0.10	0.919
ot1:colorcoding1	-1.10	0.53	-2.07	0.039
ot2:colorcoding1	0.35	0.25	1.40	0.163
condition1:ot1:colorcoding1	0.27	0.53	0.50	0.614
condition1:ot2:colorcoding1	0.23	0.25	0.91	0.365

**Table 4:** Results from the combined data from Experiments 1 and 2, overall accuracy. The predictors of condition and color-coding were effects coded with agglutinating and colour being compared to the grand average.

	Estimate	Std. Error	z value	Pr(>   z  )
(Intercept)	0.97	0.10	9.30	< 0.001
condition1	0.13	0.10	1.21	0.227
ot1	6.92	0.44	15.74	< 0.001
ot2	-2.29	0.25	-9.17	< 0.001
colorcoding1	-0.01	0.10	-0.12	0.904
condition1:ot1	0.96	0.42	2.28	0.022
condition1:ot2	0.01	0.23	0.05	0.960
condition1:colorcoding1	-0.23	0.10	-2.22	0.027
ot1:colorcoding1	-0.22	0.42	-0.53	0.599
ot2:colorcoding1	0.17	0.23	0.72	0.472
condition1:ot1:colorcoding1	-0.50	0.42	-1.19	0.233
condition1:ot2:colorcoding1	0.47	0.23	2.02	0.043

**Table 5:** Results from Experiment 3, overall accuracy. The predictors of condition and colour-coding were effects coded with agglutinating and colour being compared to the grand average.

## 2. Full model results - CLASS accuracy:

	Estimate	Std. Error	z value	Pr(>   z  )
(Intercept)	0.61	0.10	5.86	< 0.001
condition1	-0.33	0.10	-3.17	0.002
ot1	3.76	0.47	8.08	< 0.001
ot2	-2.35	0.34	-6.87	< 0.001
colorcoding1	0.05	0.10	0.45	0.653
condition1:ot1	-1.08	0.47	-2.31	0.021
condition1:ot2	0.55	0.34	1.62	0.105
condition1:colorcoding1	-0.21	0.10	-2.01	0.045
ot1:colorcoding1	-0.33	0.46	-0.70	0.482
ot2:colorcoding1	-0.08	0.34	-0.24	0.812
condition1:ot1:colorcoding1	-0.70	0.47	-1.50	0.133
condition1:ot2:colorcoding1	-0.01	0.34	-0.04	0.969

**Table 6:** Results from the combined data from Experiments 1 and 2, CLASS accuracy in the subset of the data with incorrect responses. The predictors of condition and colour-coding were effects coded with agglutinating and colour being compared to the grand average.

	Estimate	Std. Error	z value	Pr(>   z  )
(Intercept)	0.62	0.10	6.44	< 0.001
condition1	-0.35	0.10	-3.63	< 0.001
ot1	3.22	0.40	7.96	< 0.001
ot2	-2.37	0.28	-8.41	< 0.001
colorcoding1	-0.03	0.10	-0.33	0.744
condition1:ot1	-0.53	0.41	-1.32	0.187
condition1:ot2	0.62	0.28	2.22	0.026
condition1:colorcoding1	-0.13	0.10	-1.38	0.167
ot1:colorcoding1	-0.72	0.40	-1.78	0.075
ot2:colorcoding1	-0.16	0.28	-0.58	0.563
condition1:ot1:colorcoding1	-0.29	0.40	-0.72	0.471
condition1:ot2:colorcoding1	0.14	0.28	0.49	0.625

**Table 7:** Results from Experiment 3, CLASS accuracy in the subset of the data with incorrect responses. The predictors of condition and colour-coding were effects coded with agglutinating and colour being compared to the grand average.

### 3.3 Postscript

#### 3.3.1 Choice of the categories NUMBER and CLASS

Beyond the paper presented above, the reader may find the following additions helpful in contextualising the findings of the language learning experiments. To begin with, the choice of the categories NUMBER and CLASS over other options deserves further attention. Both features are common among the languages of the world; a typological survey on the use of plural markers on nouns by Haspelmath (2013) has found that over 90 per cent of the languages in the sample use nominal plural marking at least to some extent. In addition, whether a plural marker is obligatory, optional or not used at all depends in part on the ANIMACY of the noun in question and often on the distinction between animate and inanimate (Haspelmath, 2013). Similarly, the relative ANIMACY of nominal referents has a bearing on grammatical structures across the vast majority of the world's languages, whereby the scale ranges between humans and inanimates, with more or fewer subtle distinctions depending on the language in question. In addition, in some languages, the ANIMACY scales interact with other features like GENDER and DEFINITENESS (Vihman & Nelson, 2019). For instance, Differential Object Marking in Turkish has been shown to be sensitive to an interaction between ANIMACY and DEFINITENESS (Krause & von Heusinger, 2019). Studies have also shown that ANIMACY plays an important role in human cognition (Blumenthal et al., 2018; Bonin et al., 2019; Calvillo & Jackson, 2014; Kriegeskorte et al., 2008) and that it can be acquired as a grammatical feature in an artificial language learning task (Leung & Williams, 2012). At the same time, both NUMBER and CLASS can be easily depicted in images through the use of single or multiple items which fall into large nominal categories such as animal – non-animal, and these categorial differences are visually striking.

Other combinations of categories, such as CASE and NUMBER, or TENSE and ASPECT, would have been possible candidates as well, but they require more complex images. In order to illustrate CASE, one would need to depict a scene including two entities such as a subject and an object or a possessor and a possessum. Similarly, TENSE and ASPECT are marked on verbs, which require scenes rather than a static entity. In order to keep the images as simple as possible and thereby avoid potential distractions from the actual task, NUMBER and CLASS were deemed highly suitable for the research questions at hand. However, I cannot exclude the possibility that other

categories might impact participants' learning behaviour differently. It is conceivable that scenes, for instance, direct a participant's attention to a certain aspect in the image. In the experiments presented above, CLASS was more salient to the learners than NUMBER, which means that one category was noticed more or earlier than the other one. Testing different categories might cause a shift in their relative salience and affect learning in the sense that a more salient category may be learned better or faster than a less salient one.

The important point to highlight is that the pure noticing of a category in the images should be unaffected by the morphological system in question. In other words, if one category, for example TENSE, is noticed earlier than the other, for instance ASPECT, then we should observe this difference in both the agglutinating and the fusional condition, as we have seen in the experiments presented here. In the unlikely event that both categories are salient to a very similar extent, this could lead to a weaker bias for transparent structures in the fusional condition (since participants would start considering both categories at the same time). This, in turn, should mean even more similar acquisition (and error) patterns between the agglutinating and the fusional system and is not expected to produce significantly different results to those obtained in this paper. Therefore, no difference in results is expected with regards to the research questions pursued in this thesis if other categories were used, although this statement can only be verified by running further studies with different categories.

A related question that may arise is whether agglutination would facilitate learning when the visual scenes are more complex. To reiterate, the common assumption made by scholars is that agglutinating forms should be easier to learn. However, the findings presented throughout this thesis and the critical review of arguments in favour of the transparency benefit cast doubt on the idea that transparency in itself has a large impact on learnability. For this reason, learning advantages of agglutinating structures should not be assumed under all circumstances. It appears more likely that they would arise when more than two binary features are encoded in the visual scenes, where visual complexity results in a larger number of distinct meanings and distinct morphological forms. Earlier research suggests that the likelihood of learning differences increases when the numerical differences between both morphological systems increase (Brighton, 2002; Plank, 1986). However, this would constitute a considerable deviation from the experimental design used in this thesis and additional factors like frequency and the number of

morphemes would increase the differences between the two morphological systems further, which would need to be accounted for. Whether more morphological information in the visual scene more generally (through the encoding of an agent, a patient, and an action, for instance) would be learned better when the language is marking these meanings in an agglutinating manner will also need to be carefully tested, for instance by accounting for the number of morphemes in the paradigms, compared to a fusional system. Although the quantity of agglutination would be increased in longer utterances, one would need to establish which characteristics of agglutination account for the potential learnability difference.

### **3.3.2 The relationship between CLASS and NUMBER in natural languages**

A related question is how similar or different the grammatical categories of CLASS and NUMBER are in natural languages and whether their acquisition may carry intrinsic differences. Harley & Ritter (2002) describe an important distinction between both categories, in that possible feature values for NUMBER appear to be much more restricted in natural languages than feature values for CLASS: the maximum number of distinct NUMBER values found cross-linguistically appears to be four, such as singular, dual, trial and plural. CLASS, however, has been found to be more diverse. It subsumes features like ANIMACY and GENDER, whereby GENDER can have ten or more sub-classes in some languages and similarly, ANIMACY can have different dimensions and generally affects language systems in various ways (Harley & Ritter, 2002; Vihman & Nelson, 2019). The experiments presented in this thesis test CLASS distinctions based on semantically-conditioned ANIMACY, but natural languages can also have GENDER systems which are at least partly formally conditioned, such as German, French and Spanish (Corbett, 2013b). While NUMBER and CLASS vary in the complexity they can reach, both features appear related – in one way, NUMBER marking has been found to depend on ANIMACY sub-classifications (Haspelmath, 2013); in another way, ANIMACY has been described as being dependent on NUMBER under a generative approach to the organisation of pronominal paradigms (Harley & Ritter, 2002).

More precisely, in the morphosyntactic feature geometry for pronominal paradigms proposed by Harley & Ritter (2002), the representations of cognitive categories such as GENDER, NUMBER and PERSON are strongly intertwined: the participant node representing PERSON and the individuation node representing NUMBER are both directly dependent on the root node. However, the CLASS node representing

ANIMACY, GENDER and other features, is itself dependent on the individuation node - there are languages in which CLASS as a grammatical category is entirely absent, while there appear to be no languages which mark CLASS, but not NUMBER (Harley & Ritter, 2002). Furthermore, they observe that languages have been found to not contain more CLASS distinctions in the nonsingular than in the singular, which again suggests an interaction between NUMBER and CLASS.

A further difference between CLASS and NUMBER concerns their acquisition across languages. Harley and Ritter's (2002) morphological feature geometry predicts that nodes are learned prior to their dependents, which means that default values for PERSON and NUMBER should be learned by children before CLASS values in pronominal paradigms. This largely mirrors the acquisition patterns of natural languages in their sample of studies, where CLASS distinctions tend to be learned after the first PERSON and NUMBER distinctions. While NUMBER did not appear to be learned earlier than CLASS by adult participants in the experiments introduced in **Section 3.2**, the fact that CLASS is open to many more sub-distinctions than NUMBER seems to be reflected in participants' acquisition patterns, since they spend lots of attention on exploring possible CLASS categories. Note, however, that this numeric difference in possible values is not limited to the grammatical features of NUMBER and CLASS. The same imbalance also exists between NUMBER and CASE (Iggesen, 2013). In short, NUMBER and CLASS can be closely related in natural language paradigms. At the same time, the features can differ considerably in their formal complexity and acquisition order, which is mirrored in the acquisition patterns observed in the experiments presented in this chapter. As the next section will show, acquiring GENDER can be even more difficult in the presence of arbitrary links between members of individual sub-classes.

### **3.3.3 Alternative options for testing NOUN CLASS systems**

While the NOUN CLASS system tested in the experiments in this thesis is semantically conditioned, some natural languages also have partly arbitrary systems. For instance, in a typological survey of 257 languages, Corbett (2013a) finds that 84 of them have sex-based GENDER systems, 28 have non-sex-based GENDER systems and 145 languages have no GENDER system. The non-sex-based GENDER systems in the sample are in fact based on ANIMACY distinctions, but they may appear arbitrary for a

beginning learner, and exceptions exist. Similarly, also sex-based GENDER systems often contain nouns which do not fall into the GENDER category which relates to their sex (Corbett, 2013a). This can also be viewed as a form of arbitrariness from the perspective of a novel learner. Another survey by Corbett (2013b) on the same sample finds that all 112 languages in the sample that have GENDER show semantic GENDER assignment (sex-based or ANIMACY-based) to some degree, but 59 of them additionally assign nouns to GENDER based on phonological and/or morphological criteria. The latter two cases are known as formal GENDER assignment. In a small proportion of instances, GENDER assignment appears entirely arbitrary also from a linguistic perspective (Corbett, 2013b).

In principle, in the experiments presented in this thesis, testing a CLASS category which underlies a formal or arbitrary GENDER distinction instead of semantic ANIMACY would have been possible as well. However, an arbitrary GENDER system would necessarily have required the memorisation of the individual lexical items, which is why the stimuli used here were differentiated semantically instead. Since the experiments were aimed at excluding potential confounds to learning behaviour, no stem was ever repeated. This was done to avoid the possibility that some stems were more salient than others and diverting participants' attention from the suffixes. This set-up makes the testing of an arbitrary GENDER system impossible. However, beyond that, research has shown that arbitrary grammatical categories are very difficult to learn in situations of implicit learning.

For instance, Braine et al. (1990) test whether arbitrary sub-classes can be learned by children aged 7-10 years and by adults in an artificial language learning experiment. For their purpose, arbitrary is defined as the lack of an obvious link between the items in a class based on their meaning or phonological shape. In their study, nouns were inflected for three locative cases through suffixes, whereby each case had two morphemes, and their choice was exclusively conditioned on the lexical stem. For each case, one of the suffixes was used with more nouns in the language than the other suffix. Results from several acquisition tasks show that children and adults learned the sub-classes poorly and struggled to choose the correct morpheme in a generalisation task, although both morphemes were learned. Instead, both groups overgeneralised the more frequent morpheme.

Similarly, a study by Brooks et al. (1993) tested the facilitative effect of phonological similarities among members of a noun sub-class. The artificial language

was similar to that of Braine et al. (1990) and expressed three cases, each realised by two suffixing morphemes. In one condition, the nouns in each sub-class showed phonological overlap, while in another condition, nouns were assigned to sub-classes arbitrarily. Adults as well as 9-10 year-old children were able to acquire the sub-classes in a situation of implicit learning when the nouns had phonological commonalities, but not the sub-classes which were entirely arbitrary. This suggests that fully arbitrary sub-classes are unlikely to be learned implicitly in artificial language learning paradigms, which makes arbitrary GENDER class a poor alternative to the semantic ANIMACY class I tested, and fully arbitrary systems are difficult to find in natural languages in the first place (Brooks et al., 1993; Corbett, 2013b). Testing a formal GENDER class where nouns fall into sub-classes based on shared phonological properties (which are consistently shared among all members) would have been a possible alternative to conditioning ANIMACY on the semantic information in the images. According to the results from Brooks et al. (1993), this system might be learnable, but one would need to ensure that the phonological information in the noun stems does not interfere with the semantic information provided in the images.

In relation to the previous point concerning arbitrary noun classes, another theoretically possible alternative to the experimental design used here would have been to test a masculine-feminine GENDER distinction instead of the ANIMACY contrast. A study by Reeder et al. (2017) with 24 adult participants suggests that learners can acquire grammatical sub-categories in an artificial language learning task also when only distributional cues are available. In a second experiment, they find that 15 adult participants similarly learn sub-categories of the same artificial language in the presence of few exceptions to the distributional regularities. Thus, Reeder et al. (2017) are able to demonstrate that in the absence of any other cues, distributional information can be used to successfully learn arbitrary sub-categories which are similar to grammatical GENDER systems. To illustrate, in the latter, one sub-category (e.g., masculine) will often co-occur with particular articles or agreement markers while another sub-category (e.g., feminine) will occur with a different set of markers. These distributional co-occurrence patterns provide cues as to what categories exist in the language.

However, classifying noun stems into GENDER categories based on distributional information could not have been a task in the present experiments, where no stem was ever repeated for methodological reasons. Above this, each of the four

grammatical sub-categories had a single member (there was no allomorphy). Thus, the phonological information contained in the suffix combined with the semantic information from the images was sufficient to determine category membership. Using distributional instead of semantic information to determine the grammatical categories would require considerable changes to the design because more material (such as articles) would need to be added to the noun stems. This would divert the attention from comparing transparency differences between two morphological systems and make the artificial languages unnecessarily complex. The only distributional learning that may have taken place in the experiments might be the observation that the first suffix was always followed by one of two subsequent suffixes (such as an ANIMACY suffix *-ka* being followed by either *-lo* or *-mu*, depending on the number of the entities in the image). This was only possible in the agglutinating condition, but the co-occurrence patterns between two sets of suffixes could have been an indicator that the language contained two categories.

However, in principle, animate and inanimate could have been replaced by masculine and feminine GENDER. For instance, learners might be able to associate the ending *-ka* with masculine and *-mu* with feminine nouns. If the GENDER system was sex-based, then the information about the GENDER sub-categories could be retrieved from the images themselves and no additional linguistic information would be required to learn the system. However, while it is possible to use a semantic GENDER system based on masculine and feminine nouns, this was impractical to do in the experiments conducted here: since the stimuli consisted of 96 different noun stems, it would have been necessary to find masculine and feminine versions for each of them (since input was randomised across participants) and their distinctions would have to be visually clear (note that participants always saw one value per category in any one trial, so they would see a female parrot in one trial and a male lion in another, but never a male and a female parrot in the same trial). Due to these practical difficulties, the decision was made in favour of using an ANIMACY-based CLASS system.

### **3.3.4 The use of categories with equipollent markers**

Note that in the experiments presented in this thesis, each of the two categories of NUMBER and CLASS has two feature values, which are both overtly marked. This is not

necessarily the case in all feature paradigms across languages, however. Languages also display zero-marking, that is, the lack of an overt morphological form. One example is the NUMBER paradigm in English, where only the plural is marked through suffixes or irregular plurals, but the singular noun does not overtly mark singular, as in *house-houses*, *ox-oxen*. Singular forms are often left unmarked cross-linguistically, as are nominative and absolutive case (Bickel & Nichols, 2009). In Turkish, neither the nominative nor the singular are marked on nouns, while the plural and other cases are systematically marked (see example 1 in **Section 1.1**).

However, it is also curious to observe that at the same time, specifically with the feature of NUMBER, many languages also have zero plural marking, or limited plural marking. In a survey of 291 languages, Haspelmath (2013) finds that 28 languages do not mark plural on nouns at all, 133 languages mark plural in all nouns and always obligatorily, and the remaining 130 languages mark some plurals only optionally or only display plural marking on certain categories of nouns (also see **Section 3.3.1**). On the other hand, equipollent marking of singular and plural forms, that is, overt marking of both oppositions, is also found, for instance in Bantu languages (Stoll, 2015). As these examples show, NUMBER marking is variable across the languages of the world. Another interesting observation is that privative opposition (with zero exponence) has been found to occur more commonly in agglutinating systems, while equipollent marking and syncretisms are common in fusional systems (Haspelmath, 2009; Plank, 1986; Plank, 1999; Plungian, 2008). Therefore, whether overt marking of all feature values, as operationalised in the experiments at hand, is considered uncommon, depends to some extent on the linguistic context.

A question that arises concerning acquisition is whether one-to-one mapping between meaning and form can help learners acquire complex systems with zero and non-zero forms. Slobin (1973) observes that children replace zero forms by overt morphemes. However, the question could also be phrased the other way round, namely whether zero forms help the learner acquire a system with one-to-one mapping. Zero exponence and syncretism are related in that they both reduce the number of overt morphological forms in inflectional paradigms. As Plank (1986) argues, as paradigms grow large, languages tend to have other characteristics which contribute to a reduction of the number of morphemes to be remembered. In essence, both transparent mapping as well as a smaller number of forms to be learned are considered helpful for learning, although demonstrating a positive effect of the former

criterion has proven difficult. Nevertheless, the question whether zero forms are beneficial or detrimental to learning is of interest and invites further research.

What remains to be explained is why equipollent marking was chosen in the present experiments, especially since the marking of singular may not have been very familiar to the participant groups tested in this thesis (English, Turkish and Spanish native speakers). As Marsden et al. (2013) point out, a highly controlled artificial language always conflicts with the quest of understanding how natural languages are acquired. At the same time, these set-ups are important to provide insights about individual factors that contribute to learning, for which we sacrifice some naturalness in the stimuli. Furthermore, observing natural language acquisition comes with difficulties of its own, as has been shown in **Chapter 2**. To this end, rather than faithfully mimicking natural language patterns, the experimental paradigm was designed deliberately to contain two binary features with overtly marked feature values. This was done because a two-by-two paradigm (with four morphemes in total) is the only constellation where the number of morphemes is the same in an agglutinating and a fusional system (for comparison, a system of three binary features would mean six distinct agglutinating morphemes, but eight distinct fusional morphemes). This would introduce an additional confound to the study, which needed to be avoided. Testing miniature paradigms with overt markers for the singular is not novel either and has been done in earlier studies (Johnson et al., 2020; Seyfarth et al., 2014). Moreover, there is no indication in participants' learning behaviour or questionnaire responses that participants were distracted by the presence of an overt marker for the singular.

Importantly, participants were not made to believe that the artificial language they were learning was a fully-fledged language. They were told explicitly that they were learning part of a language. Theoretically, the miniature artificial language could have been a small part of a larger NUMBER-CLASS paradigm containing further feature values such as trial and human animate, some of which might display zero marking. Therefore, although overt marking of the singular may be considered untypical in natural languages, the experimental language used here is fully plausible, especially because it is not presented as a full language to participants.

### 3.3.5 Audio versions of the stimuli used in the experiments

During each of the 96 trials, participants heard the audio version of the correct word after making their selection. The stimuli, including the audio files, were adapted to participants' native language (the same applies to the experiments presented in **Chapters 4** and **5**). This was done to help participants focus on the grammatical suffixes instead of spending resources on learning novel phonetic patterns. In line with this principle, the audio stimuli for the group of English native speakers all followed the dominant stress pattern of English where stress falls on the word-initial syllable. This applied to the agglutinating and the fusional condition. For instance, words received stress in the following way: *clép-lo-pi* (agglutinating), *clép-lo* (fusional) (accents added for illustration only). Although not intended, this may have helped participants discover the word stem and potentially segment the stem from the first suffix because the word stem was auditorily salient (all stems were monosyllabic for this participant group). However, it needs to be emphasised that the stems were also visually salient because they were different in every single trial. Similarly, the fact that the same four grammatical endings were used in every trial in either condition meant that participants noticed the suffixes early on, which is also reflected in the answers they provided in the final questionnaire. Therefore, any subtle hints to word structure provided by the stress pattern that was applied was likely overridden by the visual salience of the elements in the words themselves. Not all participants reported the two agglutinating suffixes in a segmented fashion, which means that they might not have segmented them while learning the language. However, there is no indication either that the stress pattern was a hindrance to the discovery of the boundary between the grammatical suffixes.

### 3.3.6 Interpretability of the mean curves as an average across groups

One point of interest to add with regards to the figures of results shown in **Section 3.2** is that the learning trajectories of the individual participants across conditions and experiments were rather varied. This is a very general pattern I observe and it also applies to the experiments that will be presented in the next chapters (for more insights on individual learning strategies, see **Section 4.3.5**). In essence, in every condition that was run, there were some participants whose performance remained around

chance level throughout the experiment (who did not learn the grammatical system), some participants who learned all or part of the system in later stages of the experiment and some participants who learned the system very well and early on. What is important to note is that this variability in learning success is found to similar extents in the agglutinating and the fusional conditions.

In an early analysis of Experiments 1 and 2, attempts were made to split participants in each condition into three groups, namely those struggling to acquire the morphological system, those acquiring the system partially and those acquiring the system well, based on their scores throughout the experiment and their consistency of selecting correct answers towards the end. However, no obvious differences were found in the distribution of participants among these three groups between conditions. This means that although the means per condition hardly overlap with any individual's actual performance in the experiments, the means do not conceal any underlying trends. This would have been the case if, for instance, there had been a substantial number of participants in the fusional condition who were learning extremely well, suggesting that the mean learning rate in the fusional condition should actually be lower. Such a situation could have occurred if a large number of good learners had randomly all been assigned to the fusional condition. However, there is no indication of any such trend in the data or in the questionnaire responses. Therefore, the means that are shown in each of the figures above indeed represent the aggregated average per condition and are meaningful to interpret.

The same observation holds for participants' error patterns. There is individual variability in that some participants did not acquire the grammatical system well and therefore produced all three types of errors to more even proportions. Other participants learned the system partly. As described above, many participants learned CLASS before NUMBER, a pattern which tends to reveal differences between conditions regarding the extent to which it occurs. Again other participants learned the full system early on and therefore did not produce many errors of any type. **Figures 8 and 9** above indicate partially high variance on individual trials which relates to these learning differences: the extent and regularity with which CLASS correct only responses were chosen differs between participants and fluctuates across trials<sup>11</sup>. Remember that the

---

<sup>11</sup> What has been referred to as *CLASS correct only errors* and *NUMBER correct only errors* in the submitted manuscript will be referred to as *CLASS correct only responses* and *NUMBER correct only responses* in the remainder of this thesis, since these terms have been deemed more clear.

stimuli were randomised across participants and appeared at different trials and that participants who only learned the CLASS feature had a 50-50 chance of selecting the word which also expressed the correct NUMBER value. This means that they had a 50-50 chance between selecting the correct answer and selecting the CLASS correct only response. Some participants were also guessing entirely on a given trial, yet others had learned the system, but reported to have made sporadic careless mistakes. This naturally introduces variability to the error data, but again, no trends were identified which signal any differences of interest between conditions other than the stronger focus on CLASS in the fusional condition.

Finally, a large number of participants (around 50-100) was run in each experiment. This was done to reduce the impact of any outliers on the means per condition. If some participants learned outstandingly well or poorly, these instances should be distributed relatively evenly between conditions and it additionally supports the informativeness of the means.

### **3.3.7 Analysing learning rates using polynomial models**

Although the data reported in this thesis were analysed using polynomial models, it would in principle also have been possible to use generalised additive models (GAMs), which are also used to model non-linear trends in data sets. The underlying thought is that the learning curves in the data obtained here are generally bumpy – they do not increase at a steady rate, but display many small increases and decreases, while they increase overall and often flatten out towards the end. In order to capture this general trend, a polynomial model was compared to a linear model and in all cases, it fit the data better. This suggests that the increase of correct responses (and of CLASS correct only responses) over time is not linear. In comparison to polynomial models, an advantage of GAMs is that they are able to capture the non-linearity without the need of adding adequate exponent terms, which means that one does not impose a pre-specified functional form. Instead, a GAM consists of a parametric part with a linear predictor and a non-parametric part with additional functions that capture the individual bends of the non-linear curve as smoothing splines (Baayen & Linke, 2020). In this way, instead of the researcher, the model algorithm selects the form of the function (Wieling, 2018).

However, in the data obtained here, I am confident that the bumps in the curves are mostly noise in the data, for the reasons explained in the section above. Therefore, there seems to be little reason for concern that superimposing a polynomial curve misses out on meaningful variation. To the contrary, while Baayen & Linke (2020) note that polynomial models with multiple powers of  $x$  run the danger of overfitting the data, the worry with my experimental data, which I modelled with a single power of  $x$ , is that using GAMs instead would overfit the data and model variation which is not meaningful, but random. Indeed, Baayen & Linke (2020) note that in order to ensure that the fitted GAM is not too close to the data, one should favour the simplest model with as few basis functions as possible. They further describe how including random effects into the GAM reduces the wigglyness of the curve, which might otherwise occur due to overfitting. This, however, is also what the polynomial model specification that I employ does: by including item (picture) and subject as random effects, I am able to account for variation that stems from individual participants learning differently and seeing different stimuli from each other in any given trial.

While GAMs may provide a possible alternative to analysing the data at hand, there is no indication that the polynomial models used fit the data poorly. Especially since the statistical analysis reveals no difference between conditions, which was unexpected, changing the modelling approach would run the danger of finding an effect which may not actually exist in the data, especially when comparing the fitted curves visually. Since only a quadratic term (and not multiple powers of  $x$ ) is used, the danger of overfitting the data with the polynomial model I use is rather low. Second, the models employed in this thesis are not used to predict unseen data. These are two disadvantages of polynomial models described by Baayen & Linke (2020). Furthermore, polynomial models are fully appropriate for the time series data at hand (Mirman, 2017). As Wieling (2018) recommends, “if the patterns in the data (...) can be adequately represented by simple polynomials, it might be preferable to use growth curve analysis” (p.115). Therefore, the decision was made to use polynomial models here, while GAMs have been shown to have many advantages of their own due to their flexibility (Baayen & Linke, 2020; Wieling, 2018).

## Chapter 4

### Testing the effect of transparency in artificial language learning in native speakers of Turkish

#### 4.1 Introduction

Since the experiments reported in **Chapter 3** consistently failed to show an advantage of the agglutinating system, it is valid to ask whether this result at least partly stems from the fact that all participants were native speakers of English. As is well-known, the English language of the present day has lost most of its earlier inflectional system and is to a large extent isolating (Kortmann, 2005). English does not encode the feature of GENDER apart from in its pronominal system (Kortmann, 2005; Sabourin et al., 2006) and it does not differentially mark nouns based on ANIMACY. Further, NUMBER is only marked in the plural, but not in the singular. Given that neither the grammatical features in question nor excessive inflectional morphology more broadly are used in English, it is possible that these participants were struggling with the acquisition of the features of NUMBER and ANIMACY and their morphological marking on nouns. Moreover, DeKeyser (2005) observes that L2 learners with a morphologically poor first language like English may overlook morphology when it is not given additional emphasis during learning. If English native speakers were struggling with inflections on a more general level, it is possible that they found both conditions difficult, and this might have obscured any particular advantages of the agglutinating system.

This introductory section will first describe an analysis which compares the performance of English native speakers with and without knowledge of second languages, testing whether experience with morphologically richer languages can improve performance. This will be followed by a discussion of the literature on transfer effects from the L1 onto the L2 to illustrate to what extent we might expect a different learning trajectory in the experiment if we test native speakers of languages with either agglutinating or fusional morphology. I will then introduce relevant aspects of the morphology of Turkish, a highly agglutinating language. After deriving predictions about the performance of this group of participants, I will report the design and results of Experiment 4 with native speakers of Turkish.

#### **4.1.1 Potential reasons for a lack of learnability difference in an English-speaking sample**

While the experiments presented in **Chapter 3** have not revealed any overall learnability advantage of the transparent agglutinating condition, it is possible that knowledge of a second language might have helped some participants learn the artificial language better than others. Indeed, some of the English-speaking participants in Experiments 1-3 indicated to speak second languages at least at a basic level. Among the L2s listed are Spanish, French, German, Italian, Hebrew, Cantonese and Japanese, many of which are inflectionally richer than English. Thus, it is possible that some participants performed better than others in the experiments reported earlier, based on their L2 knowledge of inflectionally richer languages. These beneficial effects stemming from the L2 might have been balanced across conditions, leading to the lack of an effect observed overall, but they would be an indication that experience with certain morphological structures can be helpful for language acquisition. I therefore conducted an analysis on data from Experiments 1 and 2 comparing strictly monolingual participants to those who spoke languages besides English.

The classification of participants followed simple principles. Participants were grouped into either the strictly monolingual group or into the non-purely-monolingual group. Fully monolingual were all participants who indicated in the questionnaire that they spoke no other languages besides their native language (either via a written response or by leaving the respective field blank). All other participants were those who did indicate that they had at least basic knowledge in one or more languages besides English. Among the 180 participants tested in Experiment 1, 144 were purely monolingual and 36 had knowledge in second languages (21 in the agglutinating and 15 in the fusional condition). In Experiment 2, out of 100 participants, 81 were purely monolingual and 19 had L2 experience (7 in the agglutinating and 12 in the fusional condition). Since the second languages mostly consisted of those showing inflectional morphology to a larger extent than English, as described above, I refrained from a more fine-grained analysis. The research question to be answered was whether participants who had additional knowledge of a language other than English performed better in the experiment than those who did not.

If L2 experience improved learning in Experiments 1 and 2, the learning curves of participants with L2 knowledge should be steeper than those of monolingual participants in both conditions, and potentially steeper in the agglutinating than in the fusional condition. **Figure 1** shows mean accuracy across conditions by trial, L2 experience and colour-coding. It suggests that the learning rates between monolingual and not strictly monolingual participants were overall similar, but that participants with L2 experience learned somewhat better in the agglutinating condition. Mean accuracy across all trials in the agglutinating non-colour condition was 0.64 (SD=0.25), in the agglutinating colour condition 0.59 (SD=0.25). In the fusional non-colour condition, mean accuracy was 0.64 (SD=0.25) and in the fusional colour condition 0.59 (SD=0.23). Across the final 10 trials, mean accuracy in the agglutinating non-colour condition was 0.75 (SD=0.30), in the agglutinating colour condition 0.72 (SD=0.30). In the fusional non-colour condition, mean accuracy across the final 10 trials was 0.73 (SD=0.31), in the fusional colour condition it was 0.66 (SD=0.30).

To test the impression of slightly better learning in the group with L2 experience from **Figure 1**, the same type of statistical analysis was used as reported throughout **Chapter 3**. A logistic polynomial mixed-effects regression model was fitted to the combined data of Experiments 1 and 2 in order to test whether overall learning rates differed between conditions as a function of L2 experience and colour-coding. The learning curves were modelled using second-order orthogonal polynomials, predicting correct answer by condition, trial, L2 experience (some vs. none), colour-coding and their interactions. Since the polynomials were orthogonal, the model intercept refers to the overall average performance.<sup>12</sup> Trial was replaced by a linear and a quadratic time term; the time terms were added sequentially and the fit of the resulting two models (linear and polynomial) was compared using a Likelihood Ratio Test.

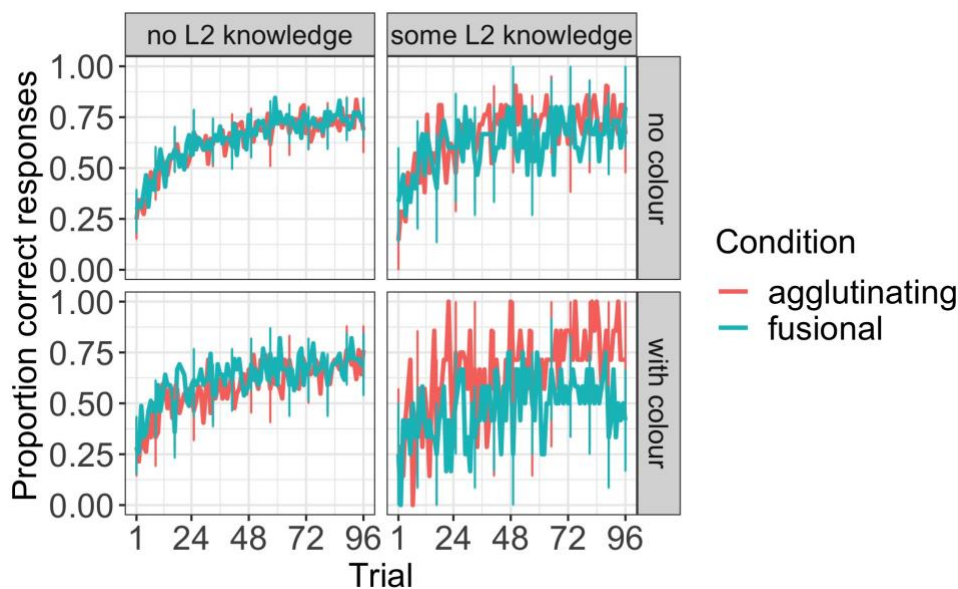
---

<sup>12</sup> All models reported in this chapter were run using the package lme4 (Bates, 2015) in R (R Core Team, 2020). The specification of the polynomial models generally follows the procedure outlined in Mirman (2014, pp.44-60). I begin with a base model including all experimental predictors as well as a linear time term ot1. A subsequent model includes the quadratic time term ot2:

Model 1: `answer_correct ~ condition * ot1 + (1 + ot1 | participant) + (1 | item)`  
 Model 2: `answer_correct ~ condition * (ot1 + ot2) + (1 + (ot1 + ot2) | participant) + (1 | item)`

Models are compared using a Likelihood Ratio Test to evaluate whether the more complex model is justified by the data. Unless otherwise noted, models included random by-participant and by-item (picture) intercepts, and by-participant slopes for the time terms. The predictors condition, L2 experience and colour-coding were effects-coded with agglutinating, some L2 experience and coloured being compared to the grand average. All models used the optimizer 'bobyqa'.

The inclusion of the quadratic time term significantly improved model fit ( $X^2=140.42$ ,  $p<0.001$ ). The polynomial model revealed a significant effect of the first time term ( $b=7.73$ ,  $SE=0.70$ ,  $p<0.001$ ) and of the second time term ( $b=-2.22$ ,  $SE=0.34$ ,  $p<0.001$ ), indicating that participants improved their accuracy over the course of the experiment and that their performance flattened out over time. There was no significant effect of condition ( $b=0.11$ ,  $SE=0.16$ ,  $p=0.494$ ), indicating that overall accuracy was similar across conditions at the midpoint of the experiment. The effect of colour-coding was marginally significant ( $b=-0.29$ ,  $SE=0.16$ ,  $p=0.064$ ), which suggests that accuracy was somewhat lower in the presence of colour cues at the midpoint of the experiment. Importantly, the effect of L2 experience was not significant ( $b=-0.07$ ,  $SE=0.15$ ,  $p=0.653$ ), indicating that participants who had experience with one or more second languages did not learn better than participants who had no L2 experience. No other effects in the model were significant, including interaction effects involving L2 knowledge (full model results can be found in **Table 2** in the Appendix to this chapter).



**Figure 1:** Mean accuracy by trial number, condition, L2 experience and colour-coding in the combined data from Experiments 1 and 2 (English native speakers). Error bars indicate 95% bootstrapped confidence intervals.

We therefore observe that participants who were familiar with other languages in the form of an L2 were not performing any differently in the learning task than strictly monolingual English speakers. This suggests that knowing a more inflectionally rich language as an L2 is not sufficient to boost acquisition of one of the inflectional types, possibly because knowledge of inflectional structures in general is not being transferred from the L2 onto the artificial language. However, this is an uncontrolled sample with participants having a range of other languages and (presumably) differing levels of expertise; it is also possible that experience with certain types of morphology through an L2 is not sufficient to facilitate learning in another L2. In this chapter and in **Chapter 5**, I therefore test whether knowledge of either agglutinating or fusional morphology, or knowledge of inflectional morphology in general, through the first language (Turkish and Spanish, respectively) improves learning in the agglutinating or the fusional condition, or both. Research on language transfer suggests that knowledge of abstract categories and morphological structures from the L1 may aid L2 acquisition. Further, native speakers of inflectionally rich languages have been shown to use decomposition more often also in second languages, which supports these learners in noticing morphological units in their L2 input. In short, there is reason to believe that the L1 matters in the L2 acquisition of inflectional morphology. I will review relevant studies on this matter in the next section before describing the morphological profile of Turkish.

#### **4.1.2 The role of the native language in the acquisition of L2 grammatical morphemes**

In **Section 2.5**, I raised the possibility that the acquisition of L2 morphology may be influenced by experience from the L1 in the broadest sense. The role of cross-linguistic influence or language transfer has been considered since the time of the ancient Greeks, but it has begun to stir more controversy around the 1960s and 1970s with the emergence of more empirical studies on language acquisition (Jarvis, 2015). In the course of the morpheme order studies, the possibility of L1 influence was considered in numerous studies, albeit with mixed results. On the one hand, the order of acquisition of English grammatical morphemes appeared to be similar across

learners with different L1 backgrounds such as Spanish, Korean, Chinese and Japanese (Dulay & Burt, 1974; Kwon, 2005). On the other hand, some researchers suggested an important role of L1 transfer (Fathman, 1975; Hakuta, 1976, as reported in Kwon, 2005). Kwon (2005) relates these differing results in part to the different theoretical stances of the researchers, who were proponents of nativist or behaviourist theories (see also Gass & Selinker, 2008). For instance, Dulay & Burt (1974) concluded that universal cognitive mechanisms were underlying the similar acquisition patterns they observed between L1 Chinese- and L1 Spanish-speaking children acquiring English morphemes. Andersen (1978), on the other hand, suggests L1 transfer as an important additional factor that needs to be considered.

Over the past decades, the role of language transfer has been continuously investigated, with research considering different aspects of language, such as vocabulary, phonology and morphology, but also transfer of semantic concepts and language processing (Gass & Selinker, 2008; Jarvis, 2015). Thereby, transfer can both support and hinder target-like language acquisition, which is captured under the notions of positive and negative transfer. Which type of transfer is happening depends on whether the learner's output is target-like or not (Gass & Selinker, 2008). While it has been found that not all characteristics of an L1 are by implication facilitative when they also appear in an L2, studies have suggested that positive transfer of certain aspects is taking place, such as transfer of word order between Spanish and Italian (Gass & Selinker, 2008). An instance of negative (i.e. interfering) transfer is the use of the English verb *become* instead of *get* by native speakers of German because *become* and the German verb *bekommen* are phonologically similar (Blumenthal-Dramé & Kortmann, 2013). I use the notion of 'transfer' to mean positive transfer in the following, since the focus of this section is whether experience with the particular kind of inflectional morphology in the L1 can facilitate learning in the experiments used in this thesis.

In this section, I will review previous findings concerning the potential role of positive transfer from the L1 onto the L2 with regards to abstract grammatical categories, experience with morphosyntax and findings concerning processing facilitation in the L2 through the L1, both offline and online. The present chapter and **Chapter 5** will complement the tests conducted with native speakers of English in **Chapter 3**. I will investigate whether the lack of inflectional morphology in the L1, as well as the type of morphology in an inflectionally richer L1, differentially influence the

acquisition of inflectional morphology in the artificial language. In this vein, I will be able to test whether morphological properties of the L1 help with the acquisition of an L2 that has similar properties. Previous research allows me to formulate predictions with regards to the different constellations of participants (L1 English, L1 Turkish, L1 Spanish) and experimental conditions (agglutinating, fusional) that I test. The predictions with regards to native speakers of Turkish will be introduced in **Section 4.1.4**, while those for native speakers of Spanish will be formulated in **Section 5.1.2**.

#### **4.1.2.1 Transfer of grammatical categories and their morphological marking**

As part of understanding second language acquisition, language transfer was frequently considered in the wake of the morpheme order studies in the 1970s as one of the determinants of acquisition order, as shown above (Andersen, 1978; Goldschneider & DeKeyser, 2001; Kwon, 2005). While the extent of its impact was debated, some researchers have found that specific English grammatical morphemes like articles and plurals are learned later if they do not exist in a learner's L1 (Kwon, 2005). A later meta-study on 18 selected studies on morpheme acquisition order conducted by Luk & Shirai (2009) found that the morphemes plural -s, articles and possessive -s can be acquired earlier or later depending on the existence of the respective grammatical categories in a learner's L1. In their review, they focus on morpheme order studies with L1 speakers of Spanish, Japanese, Korean and Chinese (for instance Hakuta, 1976; Koike, 1983; Dulay & Burt, 1974; Mace-Matluck, 1979; Pak, 1987; Rosansky, 1976). Luk & Shirai (2009) thereby test the hypothesis that structural and semantic similarity of morphemes between L1 and L2 as well as the existence of the L2 morpheme in the L1 lead to earlier learning.

Luk & Shirai (2009) show that there is a strong tendency in the combined data for L1 speakers of Japanese and Korean to acquire plural -s and articles later, and possessive -s earlier, than predicted by the natural order hypothesis as proposed by Krashen (1977). In line with DeKeyser (2005), they argue that articles are difficult for Japanese and Korean L1 speakers to acquire because the languages do not have articles. Their novelty, but also their abstractness, are argued to cause acquisition difficulty. Similarly, both languages lack obligatory plural marking, making this an area of difference between L1 and L2. In contrast, the structural similarity of possessive marking between Japanese, Korean and English is taken to lead to acquisition ease.

Data from L1 speakers of Chinese confirms the picture for plural marking, which is absent in Chinese, but results are mixed for articles and possessive -s, which they argue may in part be due to methodological issues in the original studies, such as not distinguishing between definite and indefinite articles. The learning patterns of the group of Spanish L1 speakers found by Luk & Shirai (2009) are very different from those of the other L1 groups, but can again be partly explained by language transfer. In Spanish, possession is structurally different from English in that the possessed appears before the possessor in the sentence (e.g., 'the house of my mother', Luk & Shirai, 2009, p.737). In the meta-study, possessive -s is acquired late by L1 Spanish speakers, who were also found to produce possessive constructions in English that are structurally more similar to those in Spanish. In short, Luk & Shirai's (2009) meta-analysis suggests that the existence of grammatical morphemes in both L1 and L2 facilitates acquisition, when they are semantically and structurally similar. They argue that if a grammatical feature such as plural exists in the L1, but is expressed differently from the L2, then L2 learners may be inferring the plural meaning from other sources like context, overlooking its morphological marking in the L2.

Research on the acquisition of GENDER supports Luk & Shirai's (2009) claim to some extent, since studies have shown that when GENDER exists in both L1 and L2, but the GENDER categories do not overlap, this also causes difficulty. A study by Sabourin et al. (2006) finds that L2 Dutch grammatical GENDER is learned better by those learners whose L1 marks GENDER in a morphologically similar way compared to learners in whose L1 GENDER exists, but is different from the L2. However, in both cases, GENDER acquisition was found to be more successful than when the L1 does not mark grammatical GENDER at all. For their analysis, Sabourin et al. (2006) tested native speakers of German, the Romance languages French, Italian and Spanish, as well as English. Dutch distinguishes two GENDER categories, common and neuter, which select for different definite articles and pronouns. In comparison, the German GENDER system tends to assign lexemes to the same GENDER category as in Dutch and also the agreement patterns are similar. Since morphological marking considerably overlaps between German and Dutch, this is regarded as an instance of *surface transfer*. The Romance languages selected in the study also have a binary GENDER system, but there is little overlap between the Romance languages and Dutch regarding the GENDER categories assigned to individual lexemes and the GENDER agreement system. This group of L1 speakers therefore allows the authors to test

whether abstract GENDER properties can be transferred in the absence of morphological overlap. This is regarded as an instance of *deep transfer*. Finally, English does not have grammatical GENDER aside from its pronoun system.

In two separate experiments on GENDER assignment and GENDER agreement in the noun phrase, Sabourin et al. (2006) find that GENDER assignment is learned with high accuracy in all three groups, but there are nevertheless significant differences based on the existence and similarity of grammatical GENDER between L1 and L2. The L1 German group and the group of participants speaking Romance languages as their L1 performed better in the GENDER agreement task than the L1 English group. Although the L1 German group showed a tendency to perform better than the Romance group, this difference was not significant. Sabourin et al. (2006) also find that agreement marking with middle frequency items is more affected by deep transfer than surface transfer, since the German and Romance L1 groups performed similarly well in this task. The findings are also interpreted as transfer playing a larger role in earlier stages of learning, since L1 has the largest effect on differences in outcome in this frequency range. Note that Sabourin et al.'s (2006) findings overlap with other research showing that abstract grammatical categories like grammatical GENDER, but also articles and classifiers, are difficult to acquire when they are lacking in the L1 or when their system is very different in the L1 (DeKeyser, 2005; Luk & Shirai, 2009). Finally, a study by Lowie (2000) finds with respect to derivational morphology that suffixes are learned better when there is relatively large translation equivalence (i.e. semantic overlap) between the morpheme in the L1 (Dutch) and that in the L2 (English), such as the pair *-baar/-able*. It was found that this facilitation may depend on a certain level of L2 experience. Overall, these studies indicate that conceptual and structural overlap between L1 and L2 can facilitate language acquisition.

#### **4.1.2.2 Potential transfer of morphosyntax**

The studies discussed above investigate whether abstract grammatical categories in an L2 can benefit from positive transfer from the L1. A related question is whether morphological techniques (such as agglutination and fusion) can similarly be transferred, that is, if a speaker of an agglutinating language, for example, learns agglutinating morphology in an L2 more easily than a speaker of a fusional L1. If such transfer of morphological techniques exists, this would qualify as deep transfer

according to Sabourin et al. (2006) since it is concerned with knowledge of morphological technique very generally, not with transfer of the morphological structure of a particular morpheme and the feature it expresses. An example of the latter would be plural marking through suffixes, which Sabourin et al. (2006) would regard as surface transfer. Thus, for the purpose of this thesis, we need to ask whether morphological technique can be transferred as well, that is knowledge of agglutination or fusion. Concerning the particular experiments used in this thesis, the question is whether this knowledge helps with the acquisition of separative or cumulative exponence, since I use this characteristic to contrast different levels of transparency. The question is important to raise since it is possible that the lack of much inflectional morphology in English resulted in difficulty for participants in acquiring inflections in both the agglutinating and the fusional condition in the previous experiments. Besides this, it is also conceivable that learners especially benefit from transparent agglutinating structures when this is a feature that can be transferred from their L1. Relatedly, it might be that the same type of L1 transfer happens with fusional structures, in which case non-transparent morphemes should pose less difficulty.

Contrasting the effects of an overlap of an abstract category between L1 and L2 with that of morphological richness between L1 and L2, a study by Havas et al. (2015) finds that experience with morphological decomposition from the L1 can be more beneficial for learning a morphologically marked abstract category in an L2 than experience with the same abstract category through the L1. In their study, they test native speakers of Spanish and Finnish in their learning of the GENDER (male vs. female) of animate entities marked through suffixes in an artificial language. Spanish exhibits differential marking based on GENDER, a feature that is lacking in Finnish. Further, while Spanish and Finnish are both synthetic languages, Finnish is inflectionally richer. Its nominal morphology contains 85 declensional classes and nouns can be marked for the grammatical features of NUMBER, CASE and POSSESSION, among others. Verb morphology is inflectionally even richer (Havas et al., 2015). It has been found that native speakers of Finnish use online morphological decomposition to a large extent (Havas et al., 2015; Portin et al., 2007). Further, speakers of another agglutinating and inflectionally rich language, Hungarian, have been found to also use decomposition strategies for low and medium frequency nouns in L2 Swedish (Portin et al., 2008).

In the study by Havas et al. (2015), participants completed a recognition memory task and a rule generalisation task. In both tasks, the Finnish-speaking participants performed better, which indicates better acquisition of the GENDER marking system. In line with this, they showed more awareness of the morphological rule than Spanish-speaking participants in a post-test interview. Due to the learning task employed, the results also indicate that participants were able to morphologically segment the stimuli words, with the Finnish-speaking participants being more successful in learning overall. In this way, the study by Havas et al. (2015) shows how experience with morphological decomposition can be helpful for the acquisition of new morphology, even if the underlying grammatical categories are novel. This suggests that decomposition can support the identification and potentially the memorisation of morpheme units. Even if the Spanish native speakers learned the GENDER classification system well, they may have had more difficulty retaining the appropriate morphemes. In this sense, higher morphological awareness may be linked to better acquisition of concrete grammatical morphemes. The findings by Havas et al. (2015) also suggest that processing strategies in the L1 can be beneficial for the acquisition of an artificial language, which is the method used in the experiments in this thesis.

What Havas et al. (2015) find is not that particular morphological techniques are learned better in a second language when the L1 displays the same, but that being a native speaker of an inflectionally rich L1 helps with morphological decomposition in an L2, allowing for the detection of novel grammatical features. However, in his analysis of the diachronic development of agglutinating structures within fusional systems, Igartua (2015) indirectly proposes that transparent structures are transferred onto the L2, in the sense that having an agglutinating L1 biases speakers in contact situations to make their L2 more transparent. He observes that in many languages studied in his sample, developments from fusion to agglutination happen in languages which are surrounded by languages that are strongly agglutinating, such as Cappadocian Greek dialects, Armenian, Ossetic and Georgian in the Caucasian region as well as Bengali and Marathi. His statement suggests that in these situations, learners shape their L2, but according to general morphological principles of their agglutinating L1. Igartua (2015) seemingly sees this mechanism as underlying the developments from fusion to agglutination that he observes, although he does not test his particular claim of L1 morphological influence empirically. In the experiment I report in this chapter, this test can be performed, since Turkish is a highly transparent

language. If a bias for a particular morphological structure generally exists, potentially in a way where an agglutinating structure is perceived as an abstract category, then it should mean that speakers of Turkish should learn better in the agglutinating than in the fusional condition. The studies by Havas et al. (2015) and Igartua (2015) lead to slightly different predictions. The former suggests that a morphologically rich L1 helps with the identification and learning of morphological information in the L2. The latter proposes that transparent patterns in the L1 lead to a preference for transparent morphemes in the L2.

#### **4.1.2.3 Effects of L1 processing strategies on L2 morphological learning**

While there is not much research testing the impact of a particular morphological technique in the L1 (such as agglutination) on the acquisition of the same technique in an L2, studies have been conducted which support the claim by Havas et al. (2015) that experience with morphological decomposition can benefit morphological learning in an L2. The acquisition of inflectional morphology necessarily requires the learner to recognise the morphological elements in the input (Penke, 2012). If we agree on a role of the transparency of morphemes during acquisition, we implicitly assume that a learner is aware of morphemic units which are part of the input. This may appear obvious, but there is no automatic correspondence between the two levels of analysis, at least not if one adopts a viewpoint according to which a theoretical account of language may differ from the way a learner processes language. Data-driven studies on language learning may therefore uncover learning patterns that do not necessarily overlap with a theoretical model of the language (Eddington, 2004). Put differently, there is no guarantee that learners at all levels of proficiency perceive morphemes as individual segments. This is an important point to consider, since learners in the experiments in this thesis find themselves in an early stage of acquisition. Moreover, even if morphological parsing is taking place, it may be costly to L2 learners.

This section will illustrate how research suggests that morphological processing likely differs depending on a learner's proficiency level and the structure of the L1, and moreover, that the nature of processing employed at these different stages is not undisputed. It will also highlight how morphological decomposition may relate to slower reaction times in L2 learners, suggesting it requires additional cognitive

resources. If morphological decomposition is cognitively costly, it is valid to ask whether experience with particular morphological structures from the L1 alleviates this burden. Before discussing L2 morphological processing, however, it is necessary to look at how native speakers process their language, since the two have been routinely described as different with regards to which processing route is used in which context (Clahsen, 2010; Gor, 2010). Two pathways are generally considered: full-form access of a morphologically complex word means that the word is stored and accessed as a whole. This route is considered faster, but to demand more storage space. Decomposition of a complex word into its component morphemes allows a learner to benefit from its compositional structure and saves storage space in long-term memory, but is considered slower (Portin et al., 2007). L1 and L2 speakers are understood to use both routes, but their proportion depending on language competence is still under debate (Gor, 2010; Lowie, 2000). Further, L2 online and offline processing have been shown to differ from L1 processing in different ways across a range of studies.

## **L1 processing**

Morphological decomposition in the L1 has been described as early and automatic in recent research on different languages (Gor & Jackson, 2013). It has been found that morphologically complex words (for instance words consisting of a stem and an affix, such as *teach-er*) are decomposed around 170 ms after stimulus presentation, before lexical access becomes available (Fruchter & Marantz, 2015; Neophytou et al., 2018; Rastle et al., 2004). Initial decomposition has been shown to be orthography-based and also applies to pseudo-morphological words like *corn-er*, but the processing trajectories of real and pseudo morphemes diverge over time (Fruchter et al., 2013, Fruchter & Marantz, 2015). On the other hand, a dual-route mechanism has also been proposed, according to which both full-form activation and decomposition happen simultaneously, leading to faster processing (Baayen et al., 1997). Research has also suggested that full-form access and decomposition in L1 speakers depend on the frequency of the words in question. For instance, it has been found for English and Swedish that low-frequency words are processed via the slower decomposition route, while high-frequency words are processed via full-form access, although the frequency cut-off point appears to be language-specific (Portin et al., 2007). Further, research suggests that native speakers of morphologically richer languages use

decomposition more often, that is also with L1 words of higher frequency, compared to speakers of morphologically poorer languages (Portin et al., 2007). In short, it can be assumed that L1 speakers use morphological decomposition to a considerable extent, but that it may be influenced by the morphological richness of a language.

## **L2 processing**

In comparison, research about online morphological processing in L2 learners is more sparse, although L2 learners are generally considered less sensitive to the morphological level (Clahsen et al., 2010; DeKeyser, 2005; Gor & Jackson, 2013; Kimppa et al., 2019; Jiang, 2007). However, the role of both processing routes in L2 speakers is also debated and conflicting proposals have been made. For instance, Clahsen et al. (2010) find evidence that L2 speakers rely more on lexical storage than decomposition, even at advanced proficiency levels. On the other hand, Portin et al. (2007) find the opposite, namely that L2 learners of Swedish decompose low frequency words and use full-form access only with the very high-frequency words. Both studies focused on advanced L2 learners. In the following, I will review research that suggests a development from L2-like processing to L1-like processing depending on proficiency level. What can be noted, however, is that online processing appears slower in L2 compared to L1 speakers, which indicates processing costs (Clahsen et al. 2010; Gor, 2010; Kimppa et al., 2019). It is therefore interesting to see whether the morphological type of the L1 can help with L2 processing, which should suggest easier acquisition.

Little is known about morphological processing in beginning L2 learners, but a study by Portin et al. (2007) finds that in a lexical decision task, where participants need to decide whether a string of letters is an existing word or not, two groups of L2 learners of Swedish with different levels of proficiency who were native speakers of Finnish process words similarly to native speakers: based on their analysis of reaction times and error rates, low-frequency inflected words are decomposed, while medium- and high-frequency forms appear to be accessed via the full-form route. These offline results suggest that L2 processing may become similar to L1 processing over time. However, even the less advanced group had studied Swedish for several years in high school and for 3 months at university, although reaction times and error rates were lower for the more advanced learner group. Although the authors did not test L1

transfer effects across different languages, they surmise that their L1 Finnish-speaking participants may have experienced “a positive, facilitative forward transfer effect of the overall morphological skills from the native language to the L2”, since Finnish is very rich in inflectional morphology (Portin et al., 2007, p.150). This might accelerate the process of acquiring L1-like processing in an L2.

The question of whether L1 processing strategies can be transferred to an L2 is followed up by another study by Portin et al. (2008). They compared native speakers of Hungarian (agglutinating) and Chinese (isolating) in a lexical decision experiment on Swedish inflected nouns. It was found that L1 speakers of Hungarian employed morphological decomposition in low and medium frequency Swedish words and full-form access in high frequency Swedish words. In comparison, the group of L1 Chinese speakers used full-form processing throughout, independent of noun frequency. Since Hungarian is inflectionally rich, native speakers are thought to resort more often to morphological decomposition and to transfer this strategy to processing of an L2. Similarly, the L1 Chinese group is understood to transfer their processing strategy (full-form access) onto the L2. In other words, processing strategies in general appear to be transferred from the L1 onto the L2. An interesting aspect is added to the study in that the target language, Swedish, is fusional, but the results show that learners employ native-like processing strategies independent of the L2 they are learning. Like Chinese, English morphology is relatively isolating, which may suggest that native speakers of English were decomposing the stimuli words less than native speakers of Turkish and Spanish will in the experiments I present in this chapter and in **Chapter 5**. We can therefore tentatively assume that overall accuracy rates will be higher in the experiments with Turkish and Spanish native speakers and that they may show better acquisition in the agglutinating condition, where morphological segmentation is key to benefitting from morpheme transparency.

Further studies complement the view that second language learners can attain native-like processing strategies. For instance, research by Kimppa et al. (2019) also has shown that with increased L2 proficiency in Finnish, learners become more similar to L1 speakers in their processing of inflectional, and also derivational, morphology. They employed a passive listening paradigm with low task demands. Overall, their results suggest that L2 learners’ sensitivity to morpheme units is initially weak, but that learners begin to process inflected words more similarly to native speakers with increasing proficiency, drawing more on early automatic morphological parsing than

on full-form access in advanced stages, whereas late morphological parsing effort (around 170-240 msec) was detected in beginning learners. Gor & Jackson (2013) find that L1 English speakers acquiring Russian at different proficiency levels all use morphological decomposition with regular verbs, while the processing route for semi-regular and irregular verbs is dependent on proficiency. The more advanced their level of Russian, the more native-like the learners' processing route was found to be. Gor & Jackson (2013) further suggest that L2 morphological processing is influenced by L1 transfer, in that morphologically richer L1s will lead to more decomposition in the L2. In their study, they tested native speakers of English, an inflectionally poor language, and expect that L1 speakers of inflectionally richer languages will show even more similar processing strategies to native speakers, which under their account means more morphological decomposition. Again, with respect to the present experiment, this suggests that native speakers of Turkish should use more or earlier morphological decomposition than native speakers of English. This might mean that decomposition is at play even at the early stage of learning tested, and could result in higher accuracy rates for the L1 Turkish group.

A study by Jacob et al. (2019) also found that heritage speakers of Turkish use early form-based decomposition of morphologically complex Turkish words, similar to native speakers, in a masked priming study. A difference they observe between both groups is that heritage speakers are slower, which again points to increased processing effort. However, their findings support the view that with increased proficiency, native-like processing including decomposition is possible. Note that Jacob et al. (2019) did not differentiate words based on frequency counts in their analysis. Although investigating processing in early bilinguals, two studies on Finnish-Swedish bilinguals found that in either language, bilinguals employed more decomposition than full-form access across different word frequencies compared to monolinguals (Lehtonen & Laine, 2003; Lehtonen et al., 2006). This was interpreted as an effect of bilinguals' reduced exposure to the words in either language: since less exposure reduces the frequency of words for the speakers in question, this may have resulted in the change of processing strategy in bilinguals compared to monolinguals, namely to decomposition, which has been found to be more prevalent with lower frequency words (Portin et al., 2008). Taking the findings from these studies together, inflectional richness of the L1 might interact with word frequency in that L1 speakers of inflectionally rich L1s may be relying more on decomposition in an L2 due to their

L1 morphology and reduced exposure. In turn, it might mean that Turkish native speakers in the present artificial language learning experiment could be expected to employ decomposition strategies, first because of their L1 processing, and second potentially also because the stimuli words have been observed with very low frequency at learning onset. This may contribute another factor for increased decomposition.

Finally, a study by Bosch et al. (2017) found that while advanced late L1 Russian learners of L2 German were native-like in their offline processing of German adjectives, the online processing profile of the L2 learners differed from that of native speakers. This finding suggests that even if online processing strategies between L1 and L2 speakers differ in certain contexts, L2 learners can still develop native-like representations. Across the studies shown above, it has been suggested that non-native speakers develop native-like processing strategies with increasing proficiency and that this happens faster in offline contexts (Gor & Jackson, 2013; Portin et al., 2007; Portin et al., 2008) than in online contexts (Kimppa et al., 2019; Bosch et al., 2017). For the present experiment, this means that even if Turkish native speakers do not benefit from their L1 online processing strategies, their morphological awareness may still help them in the offline learning task.

## **Summary**

In summary, research on language transfer has investigated several areas of language structure that are relevant for the present analysis. First, it is possible that the existence of L2 abstract grammatical features in the L1 has an impact on how the artificial language will be learned by native speakers of Turkish. Studies have found that when morphemes show structural and semantic similarity between L1 and L2, such as plural marking through nominal suffixes, they are acquired more easily (Luk & Shirai, 2009; Sabourin et al., 2006). In contrast, abstract features that are lacking in the L1 can cause acquisition difficulty (DeKeyser 2005; Luk & Shirai, 2009). Another possible area of L1 transfer is experience with a particular morphological technique, in this case especially separative or cumulative exponence. Although Igartua (2015) indirectly suggests that when L2 learners are native speakers of an agglutinating language, they impose agglutinating features onto their L2's morphology, the influence is likely less direct. Havas et al. (2015) have shown that experience with morphological

decomposition through the L1 can aid in the acquisition of L2 morphology, more specifically, the acquisition of an artificial language. This effect on learning was also stronger than L1 transfer of the abstract grammatical feature in question. In addition, it is conceivable that experience with agglutinating transparent structures will instill an expectation in participants that morphemes are generally transparent.

It has further been found that L2 processing is malleable, on the one hand depending on proficiency. On the other hand, it has been found that L1 speakers of morphologically rich languages use decomposition with a larger amount of words in the L1 and that processing strategies from the L1 can be transferred onto the L2 (Portin et al., 2007; Portin et al., 2008). It is possible that this processing technique and the resulting morphological awareness helped learners in the study by Havas et al. (2015) identify and retain grammatical morphemes more easily. Given that morphological processing in the L2 is considered costly, it is intriguing to see that morphological patterns in the L1 can increase the use of this processing strategy, which may ultimately also impact online processing (Clahsen et al., 2010; Gor, 2010; Kimppa et al., 2019). In addition, research suggests that L2 learners use native-like processing and decomposition more readily in offline tasks, which is the type of task used in the present artificial language learning experiment (Bosch et al., 2017; Portin et al., 2007).

Research by Lehtonen & Laine (2003) and Lehtonen et al. (2006) further suggests that a lower amount of exposure to morphologically complex words in a language reduces their frequency for a speaker, leading to more decomposition. Overall, it needs to be noted that controversy exists as to how processing routes depend on the type of morphology (regular/irregular, L1/L2), the language in question and word frequency, among others (Gor, 2010; Gor & Jackson, 2013). However, findings from the studies reviewed here suggest that it may be possible that the combination of L1 Turkish speakers with the low frequency of artificial words tested, due to the early stage of learning, may increase the amount of morphological processing in this group of participants. In turn, morphological decomposition should help with the identification of the grammatical morphemes and may support acquisition of the underlying grammatical categories. The subsequent section will illustrate some relevant aspects of Turkish grammar before **Section 4.1.4** will add the observations from the literature to derive predictions about the learning patterns of the L1 Turkish group in the artificial language learning experiment.

### 4.1.3 Relevant aspects of Turkish morphology and grammar

#### 4.1.3.1 Turkish as a synthetic agglutinating language

Turkish is often regarded as a textbook example of an agglutinating language (Bickel & Nichols, 2013a; Haspelmath, 2009; Igartua, 2015; Kortmann, 2005; Laaha & Gillis, 2007). Its morphology is highly regular and it displays many of the prototypical characteristics of agglutinating morphology introduced in **Section 1.2**. It dominantly displays separative exponence, very few inflectional classes, lack of suppletion, uniformity of grammatical categories and relatively long words (Acarlar and Johnston, 2011; Dressler et al., 1987; Dressler et al., 2006; Hengeveld & Leufkens, 2018; Plank, 1999). Needless to say, Turkish morphemes also display a high degree of transparency (Acarlar and Johnston, 2011; Dressler et al., 1987; Hengeveld & Leufkens, 2018). For instance, an analysis by Laaha & Gillis (2007) finds that around 90% of Turkish noun forms and over 80% of verb forms are transparent. Despite its highly regular morphology, vowel harmony and consonant (de)voicing cause some variation in morphological realisations found in the language (Jacob et al., 2019).

Turkish is also highly synthetic and predominantly uses suffixation (Bickel & Nichols, 2013c; Jacob et al., 2019). That Turkish is very rich in inflectional morphology is demonstrated by the fact that it marks 15 inflectional categories on nouns and 37 on verbs (Laaha & Gillis, 2007). The authors further find that in Turkish, 30% of verbs have one suffix, 40% have two suffixes and 10% have three suffixes. A few per cent of Turkish verbs are marked with four suffixes, while the remaining almost 20% of Turkish verbs are suffixless (Laaha & Gillis, 2007). Despite the large number of inflectional forms, Turkish is highly regular in many respects: it does not have any inflectional classes and little allomorphy aside from phonological conditioning. For instance, the plural marker only has two allomorphs, *-ler* and *-lar*. Although vowels change to obey Turkish vowel harmony, it follows regular patterns, which, together with the little amount of variation of forms, makes Turkish word forms highly predictable.

An example of transparent and regular NUMBER-CASE marking on Turkish nouns is provided in Comrie (1981). Turkish marks two NUMBERS (singular and plural) and six CASES (nominative, accusative, genitive, dative, locative and ablative). While singular and nominative are not marked overtly (zero morphemes), all other forms are fully regular, i.e. invariant in their shape, and correspond to those shown in (1). As Laaha

& Gillis (2007) note, the Turkish “case system consists of a single paradigm that is fully productive and regular” (p.47). The words are also fully compositional and segmentable into stem, NUMBER suffix and CASE suffix:

(1) Declension of the Turkish noun *adam* ('man')

	Singular	Plural	
Nominative	<i>adam</i>	<i>adam-lar</i>	
Accusative	<i>adam-ı</i>	<i>adam-lar-ı</i>	
Genitive	<i>adam-ın</i>	<i>adam-lar-ın</i>	
Dative	<i>adam-a</i>	<i>adam-lar-a</i>	
Locative	<i>adam-da</i>	<i>adam-lar-da</i>	
Ablative	<i>adam-dan</i>	<i>adam-lar-dan</i>	(adapted from Comrie 1981)

Note that the Turkish NUMBER-CASE system stands out typologically, since it has been observed that the two features are expressed cumulatively in many languages (Bickel & Nichols, 2013b; Plungian, 2008).

The following examples illustrate the morphological richness of the Turkish language. Affixes occur in the form of suffixes, which are highly transparent. However, almost all PERSON-NUMBER suffixes are cumulative exponents.

(2) *Haber-ler-i komşu-lar-dan duy-du-m.*  
 news-PL-ACC neighbour-PL-ABL hear-PAST-1<sup>st</sup>SG  
 I heard the news from the neighbours.

(3) *Doktor-u ara-(y)a-ma-(y)acak-sın.*  
 Doctor-ACC call-(ability to call)-NEG-FUTURE-2<sup>nd</sup>SG  
 You will not be able to call the doctor.

(4) *Sepet-te-ki-ler-i ev-e götür-dü-ler.*  
 basket-LOC-DEFINITE-PL-ACC house-DAT take-PAST-3<sup>rd</sup>PL  
 They took the things (ones) in the basket home.

(Examples are from personal correspondence with one of the Turkish language informants)

With regards to the regularity of Turkish morphology, Acarlar and Johnston (2011) find that Turkish-learning children with atypical language development show a very similar learning rate of Turkish verbal suffixes as age-matched as well as mean-length-of-utterance-matched children. This ties in with the overall picture that Turkish-learning children have less difficulty acquiring the morphological system of their L1 than

children learning other Indo-European languages (Aksu-Koç & Slobin, 1985; De Houwer and Gillis, 1998; Hengeveld & Leufkens, 2018; Laaha & Gillis, 2007). In other words, Turkish morphology has many features that appear to support learnability.

#### 4.1.3.2 ANIMACY-based Differential Object Marking in Turkish

Although Turkish does not have any noun classes and does not mark grammatical GENDER, it displays a distinction based on ANIMACY in the area of Differential Object Marking (DOM) (Aksu-Koç & Slobin, 1985; Chamorro et al., 2016; Krause & von Heusinger, 2019). This latter aspect of grammar refers to “the optional marking of the direct object” through accusative case (Krause & von Heusinger, 2019, p.171). However, the role of ANIMACY in Turkish is limited, mirrored in the fact that it has long been believed that DOM in Turkish is solely dependent on DEFINITENESS. However, Krause & von Heusinger (2019) present evidence from acceptability judgment tasks which suggests that native speakers use DOM in relation to an ANIMACY scale. In Turkish, whether the direct object is marked or not depends on several factors, including DEFINITENESS, SPECIFICITY and ANIMACY (Krause & von Heusinger, 2019). For instance, all definite direct objects are marked overtly with the accusative suffix *-(y)I* (underlying variation due to vowel harmony), as shown in (5) below.

- (5) Zeynep          adam-ı/\*-Ø                  gör-dü.  
 Zeynep          the man-ACC/\*-Ø.          definite animate NP  
 ‘Zeynep saw the man’.

*(adapted from Krause & von Heusinger, 2019)*

ANIMACY, on the other hand, decides whether indefinite direct objects can move to sentence-initial position without additional modification: they can if they are animate entities. The animate interrogative pronoun *kim* (who) needs to receive accusative marking when used as a direct object, while the inanimate interrogative pronoun *ne* (what) does not obligatorily receive accusative marking, as shown in (6).

(6) a) Hasan                      kim-i/\*kim-Ø                      gör-dü?  
           Hasan                      who-ACC/\*who-Ø                      see-PST.3SG  
           Whom did Hasan see?

b)     Hasan                      ne-yi/ne-Ø                      gör-dü?  
           Hasan                      what-ACC/what-Ø                      see-PST.3SG  
           What did Hasan see?

*(adapted from Krause & von Heusinger, 2019; original examples taken from Comrie, 1975)*

ANIMACY is also scalar in Turkish, evidenced in the observation that plural animals are less likely to take plural number agreement on verbs compared to plural entities referring to humans (Krause & von Heusinger, 2019). The results from acceptability judgment ratings conducted by Krause & von Heusinger (2019) revealed that DOM was differentially affected by ANIMACY categories (human, animal, inanimate) when other factors like SPECIFICITY were controlled for. Thus, while ANIMACY is not marked overtly through noun class, for instance, it does play a role in Turkish grammar and Turkish native speakers appear to be sensitive to various ANIMACY distinctions.

For the purpose of the experiments conducted in this thesis, native speakers of Turkish are an ideal candidate participant group. If L2 learners can have the kind of experience with transparent morphology which leaves an imprint on their L2 acquisition strategies, then native speakers of Turkish should have it. Turkish is an almost prototypical example of the agglutinating type, which means that L1 speakers of Turkish are very familiar with agglutinating-type structures, including morpheme transparency. Jacob et al. (2019) also note that the high level of “productivity of Turkish morphology has been claimed to lead to the prevalence of combinatorial processing and the decomposition of complex word forms into their morphological constituents during word recognition” (p.178). For these reasons, the language was chosen in order to test the hypothesis that experience with agglutinating inflection can have a positive effect on learning transparent morphology in an L2. Whether it is through transfer of agglutinating structures or through a predisposition to decompose morphologically complex words to a larger extent, which makes morpheme units more noticeable, L1

speakers of Turkish meet these theoretical preconditions. It needs to be mentioned, however, that the present study is not able to confirm at which level such benefits may be operational, since it looks at the end-product of learning and not at morphological processing directly. This chapter tests whether transparency in the L1 exerts a bias for or relates to learning ease with transparent morphology in the L2.

#### **4.1.4 Predictions about learning outcomes for the L1 Turkish group considering language transfer**

Having described morphological aspects of the Turkish language, we are now in a position to formulate predictions for the learning performance of the L1 Turkish participants. Their performance can also be compared to that of the English native speakers in Experiment 1, since both experiments use the same methodology, as we will see in **Section 4.2**. The artificial language used in the experiments in this thesis has two grammatical features, NUMBER and CLASS (ANIMACY). Concerning the former, English and Turkish are similar in that they mark plural through suffixation on nouns, but do not mark singular, while the artificial language marks both as suffixes. However, the structural and semantic similarity of the morphemes between the L1s English and Turkish and the artificial L2 is still high and the feature is familiar to participants from their L1. We should therefore not expect major acquisition difficulty of the NUMBER feature (see e.g., Luk & Shirai, 2009). However, as the experiments with native speakers of English have shown, NUMBER appears more difficult to learn than CLASS, which may have to do with learning biases in adults and the fixation of human learners on ANIMACY (Blumenthal et al., 2018; Bonin et al., 2019; Calvillo & Jackson, 2013; Kriegeskorte et al., 2008; Strickland et al., 2017). Since English and Turkish show a similar marking for NUMBER, I expect native speakers of Turkish to perform similarly to L1 English speakers with regards to this feature. However, the picture looks different for ANIMACY, which is not consistently marked in either English or Turkish. Additionally, there are slight differences between the L1s. While English hardly displays any grammatical distinctions based on ANIMACY (apart from genitive marking and the pronoun system), Turkish uses DOM which has been shown to be partly dependent on ANIMACY. However, like English, Turkish does not mark NOUN CLASS through GENDER. The difference between Turkish and English, which appears small, is not expected to cause considerably different outcomes in the L1 Turkish group regarding

the two grammatical features. However, having more experience with transparent morphology may lead Turkish-speaking participants to produce more CLASS correct only responses in the fusional condition than native speakers of English did. Given the findings stated above, I expect L1 Turkish speakers to learn CLASS better than NUMBER, which should lead to incorrect choices of words in which only CLASS is correct, to at least the same or a larger extent as observed for the L1 English group.

**Section 4.1.2.3** has shown that both full-form access and decomposition of morphologically complex words are employed by L2 speakers, but their use may depend heavily on context. Decomposition has been described as slower than full-form retrieval (Portin et al., 2007), but at the same time, researchers have noted that recognition of morphemes is the first step in their acquisition (Dressler, 2005; Penke, 2012). Taking the issue of processing speed aside, early decomposition is certainly beneficial to a learner, whether offline or online. In the same vein, as described in **Chapter 2**, the transparency of agglutinating morphemes can only be exploited by a learner if these are also recognised as units in the input. A learner will therefore need to segment complex words into their morphemic elements. If these words were only accessed as wholes, transparency would not be expected to play a role during acquisition. In light of this, experience with morphological decomposition from the L1 would certainly be an asset during the L2 learning process, if one assumes that positive transfer is taking place in this realm.

Given the research findings illustrated in **Section 4.1.2.3**, it can be assumed that having an inflectionally rich L1 will increase the extent of morphological processing employed by the L1 Turkish participants. Not much is known about processing strategies at the earliest stage of L2 learning, but the study by Havas et al. (2015) suggests that L1 processing strategies can support learning of an artificial language. Research has also suggested that less frequent words will be decomposed more readily by L2 learners, which may also apply to the early stage of L2 acquisition tested in the current experiment (Lehtonen & Laine, 2003; Lehtonen et al., 2006; Portin et al.; 2008). Experiments 1 and 2 from **Chapter 3** have provided some evidence that offline segmentation at the earliest stages of acquiring L2 inflectional morphology is successful and a considerable number of English-speaking participants reported the agglutinating morphemes in a segmented fashion. Considering that the participants in these experiments were native speakers of a largely isolating language, this might suggest that decomposition is relatively robust early on. However, the experiments did

not test online processing, which could come with additional processing costs. Moreover, the transparent agglutinating system was not learned better than the fusional system by English native speakers, which could indeed be due to processing difficulties, given that English does not contain many inflectionally complex words. In other words, segmentation of grammatical morphemes in both conditions may have happened successfully, but it might have been costly. This might have prevented English-speaking participants from exploiting the transparency advantage.

Given that native speakers of Turkish are more experienced with decomposition from their L1, they are expected to transfer their processing strategies from the L1 and to have greater morphological awareness (see e.g., Gor & Jackson, 2013; Portin et al., 2008). Whether they decompose the stimuli words online or not, this awareness is predicted to lead to easy segmentation of the grammatical morphemes and consequently, to better acquisition of the grammatical features. This is expected to result in higher overall accuracy rates in both conditions compared to the results from Experiment 1 with native speakers of English. This effect may additionally be supported by the fact that Turkish is an agglutinating language and that this exerts a preference for agglutinating structures, as Igartua (2015) suggests. Note, however, that this effect is considered less plausible than the transfer of processing strategies.

With regards to the question of processing differences between the agglutinating and the fusional condition, another prediction can be made. Agglutinating words typically contain several morphemes due to the compositional structure of agglutinating systems. This is reflected in the artificial words as well and means that agglutinating words contain more morpheme boundaries. It is possible that this added processing difficulty for English-speaking participants. However, this level of decomposition should be familiar to native speakers of Turkish. The design of the experimental stimuli with highly variable stems should have supported segmentation additionally. Ultimately, it is assumed that the constellation of the L1 Turkish being inflectionally rich, an agglutinating language, the variability of the artificial stems and the words being relatively infrequent for the learners in the experiment, due to their novelty, results in this participant group being most likely to be able to exploit the benefits of transparent morphology. It is therefore predicted that native speakers of Turkish will learn better in the agglutinating compared to the fusional condition. It is also expected that native speakers of Turkish will benefit most from the agglutinating

condition and show the largest difference in learning between agglutinating and fusional morphology of all L1 groups tested in this thesis. I will return to this prediction in **Section 5.4**.

In short, the L1 Turkish group is expected to perform similarly to the L1 English group in Experiment 1 with regards to acquiring the features NUMBER and CLASS. The latter is again expected to be learned better, resulting in CLASS correct only responses to be the most frequent error type. Due to their experience with transparent morphology, the L1 Turkish group might even show a slightly stronger focus on CLASS in the fusional condition than the L1 English group. Native speakers of Turkish are also expected to show higher overall learning accuracy in both conditions than native speakers of English. Finally, the L1 Turkish participant group is expected to learn significantly better in the agglutinating compared to the fusional condition.

## **4.2 Methods**

The experiment conducted with native speakers of Turkish (Experiment 4) is overall the same as Experiment 1 in **Chapter 3**. However, the stimuli and instructions were fully adapted to the Turkish language to ensure similar testing conditions.

### **4.2.1 Participants**

Different from the experiments presented in **Chapters 3** and **5**, participants were recruited via Prolific instead of Amazon Mechanical Turk. This was done because the available number of workers on Amazon Mechanical Turk who were situated in Turkey was too low to guarantee a timely completion of the experiment, and it was not possible to specifically target workers in countries other than Turkey based on their native language. On Prolific, it is possible to only allow those participants who had indicated to be native speakers of Turkish on their profile. I made use of this filter, but it was not possible to select only participants based in Turkey because the numbers were too low. Since the switch to Prolific meant adaptation of the experimental coding framework, this experiment was conducted after the experiment described in **Chapter 5**, but it is reported first due to the underlying theoretical considerations. There is also a potential for the performance to differ between English and Turkish native speakers due to the change of recruiting platform, and research suggests that the participants available on both crowdsourcing sites differ in significant ways (Litman et al., 2021;

Pe'er et al., 2021). If performance is indeed different between the L1 English and the L1 Turkish participants, it is possible that this stems in part from the different testing platforms. However, it will still be possible to evaluate Turkish L1 speakers' success in acquiring either the agglutinating or the fusional condition, and compare these.

64 participants were recruited online via Prolific, who all self-reported to be native speakers of Turkish and also self-reported to be at least 18 years old. The country of residence of the participants is not known for this experiment. Participants were each paid £4 for their time. They were randomly assigned to one of the two conditions (33 to the agglutinating condition and 31 to the fusional condition)<sup>13</sup>. I used participants' Prolific ID and Prolific's completion code system to ensure that participants only took part in the experiment once.

It is also noteworthy that while it was possible to find 64 eligible participants via Prolific over the course of 11 weeks, it took increasingly longer to find these participants compared to the L1 English speaking group on Amazon Mechanical Turk. Additionally, a much larger proportion of those who participated in the experiment indicated in the questionnaire that they had taken notes, which led to the exclusion of their data and considerably reduced the amount of data that could be used for the analysis. Finally, the number of native speakers of Turkish on Prolific is limited. For these reasons, the final participant number is lower than the sample size in the L1 English experiments, and also the experiment presented in **Chapter 5**.

## **4.2.2 Creation of stimuli and additional experimental material**

### **4.2.2.1 Artificial stems**

As in the experiments reported in **Chapter 3**, the artificial language consisted of 96 nouns which referred to either everyday objects such as household items and clothing

---

<sup>13</sup> 43 additional participants took notes during the experiment, based on their answers in the questionnaire following the experiment; their data was excluded. 14 additional participants indicated in the questionnaire that they were bilingual. These participants were also excluded since the number was too low to be entered as a separate group into the analysis. As in the experiments presented in **Chapter 3**, bilingual was defined as speaking more than one language from birth or from a young age, and being fluent in it. The purpose was to exclude participants who had a very high level of proficiency in an additional language from early in life. The remaining data from 64 participants was used for the analysis. Of these, 63 participants indicated having at least basic skills in at least one other language apart from English. The most frequently mentioned L2 was by far English, but also German, Spanish, Italian, French, Dutch, Russian and Portuguese were indicated by a various participants.

(inanimates, 48 items) or to animals (animates, 48 items). The language again had four suffixes, encoding the binary features NOUN CLASS (animate or inanimate) and NUMBER (singular or plural). This ensured that the only difference between the present and the previous experiments was the participant group that was tested. However, the artificial stems were adapted to the new participant group. The stimuli stems in the experiments presented in **Chapter 3** were all monosyllabic pseudo-English stems which adhered to English phonotactics. This was done in order to remove any additional burden that might arise from learning novel phonotactics, so that participants could fully focus on learning the grammatical markers instead. For the present experiment with Turkish participants, the same procedure was repeated.

Stems were made to follow Turkish phonotactics, for the same reason stated above, namely to avoid participants spending extra effort on learning novel phonotactic patterns. In order to create the Turkish stimuli, a native speaker of Turkish, who was a student at the University of Edinburgh, was tasked with the creation of pseudo-Turkish stimuli under my supervision<sup>14</sup>. We began with the 96 pseudo-English stems and changed them in minimal ways into stems following Turkish phonotactics. We thereby paid attention to the following points: pseudo-Turkish stems were again monosyllabic and we ensured that they were not existing words in Turkish. Similarly, to the best of our knowledge, we excluded any potential stems which would yield an existing Turkish word through the affixation of the fusional suffixes *-mu*, *-ka*, *-pi*, *-lo* or of the agglutinating grammatical endings *-muka*, *-mulo*, *-mupi*, *-kamu*, *-kapi*, *-kalo*, *-pimu*, *-pika*, *-pilo*, *-lomu*, *-loka*, *-lopi*. We further ensured that stems were not too similar to each other and that a variety of letters from the Turkish alphabet were used. Examples of pseudo-Turkish stems are: *cez*, *gim*, *klep*, *nüz* and *surç*.

In addition to the novel stems, the corresponding audio for the stems was updated as well. These were adapted to Turkish pronunciation using the Mac speech synthesizer with voice Yelda.

Morphemes used in the agglutinating and the fusional condition were again identical and remained unaltered from Experiment 1 in the English version: *-mu*, *-ka*, *-pi*, *-lo*. Similarly, the transparency of the form-meaning mappings of these morphemes was manipulated according to condition, as shown in **Table 1**. Mappings

---

<sup>14</sup> I would like to thank the four native speakers of Turkish who were involved in the preparation of this experiment: Deniz Özyıldız, Efe Akgul, Enes Aydoğan and Özge Spike.

between morphemes and grammatical meanings were again randomised across participants in each condition as in Experiment 1.

	animate singular	animate plural	inanimate singular	inanimate plural
agglutinating	zirt-mu-pi	zirt-mu-lo	zirt-ka-pi	zirt-ka-lo
fusional	zirt-mu	zirt-ka	zirt-pi	zirt-lo

**Table 1.** Illustration of the mapping of grammatical meanings onto suffixes in both conditions. *zirt-* is the noun stem in each case, hyphens are provided for illustration only and did not appear in the stimuli.

#### 4.2.2.2 Translations and creation of additional material

Besides the stimuli used during trials, the study information form which participants needed to accept in order to be able to participate was also translated into Turkish with the help of native speakers. Similarly, all instructions that participants received during the experiment were translated into Turkish to avoid any interference of a third language (English) between the mother tongue and the artificial language to be learned. The final questionnaire was also entirely translated into Turkish.

Finally, I introduced four tests which were meant to detect any non-native speakers. During the experiments reported in **Chapter 3**, it became clear that some participants were not native English speakers, based on their answers in the final questionnaire. This was despite the fact that participants needed to confirm that they were native speakers of English prior to the experiment. To circumvent the problem of having to exclude data from non-native speakers again, I employed two tests prior to the experiment and two tests between the experiment and the final questionnaire (the tests can be found in the Appendix to the thesis). These tests were forced choice questions with four options and participants needed to select the option that was grammatically correct in Turkish. One of them tested knowledge of an idiom. Together with the Turkish informants, I chose the forced choice options to contain small grammatical errors which are still common among advanced learners and where googling translations of the options would not provide any cues to any ungrammaticality. The rationale behind this was that finding the correct answer would

be laborious enough for non-native speakers to either select randomly or to discontinue the experiment voluntarily. Since I used four tests with four options each, it was highly unlikely that a participant would provide all four correct answers based on guessing. If one or more answers were incorrect, the participant's data was excluded.

After completion of the full experiment, one of the Turkish informants was tasked with the translation of the questionnaire responses from Turkish into English, which was necessary to be able to exclude all participants who did not meet the participation criteria (taking notes, etc.) and to ensure full understanding of participants' strategies.

### **4.2.3 Procedure**

The procedure was the same as in the experiments presented in **Chapter 3. Figure 2** shows a trial in the agglutinating and a trial in the fusional condition, each consisting of an image and a choice of four words that could describe the image. Participants again completed 96 trials which each showed a unique picture to avoid any image or stem repetition. In addition to making the stems pseudo-Turkish, this removed participants' attention from learning the mapping between images and stems and directed it towards learning the inflectional morphemes and their underlying meaning with regards to NOUN CLASS and NUMBER. The combinations of grammatical meanings (animate+singular, animate+plural, inanimate+singular, inanimate+plural) were again balanced across the 96 trials. Directly after the experiment, participants completed a final questionnaire which was equivalent to the one used with the native English-speaking participants.

If a participant failed to provide the correct answer in one of the two native speaker tests prior to the experiment, a message appeared on the screen saying that the selected answer was incorrect and that the participant was not able to continue the experiment. If an incorrect answer was selected in the final two tests, no message appeared and the participant was able to complete the experiment, but their data was excluded. The experiment lasted for approximately 25 minutes.



**Figure 2:** Example trials in the agglutinating condition (left) and in the fusional condition (right) in Experiment 4, showing an animate singular trial and an inanimate plural trial, respectively.

## 4.3 Results

### 4.3.1 Overall accuracy

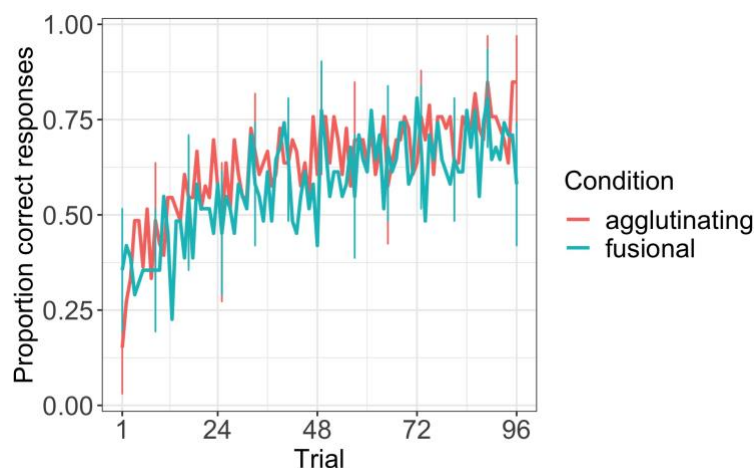
Similar to the previous experiments, this experiment was designed to compare performance across conditions over time. The adaptive tracking paradigm by default required participants to take a guess initially, and I expect performance to be similar across conditions early on. However, there is a potential for learning rates to diverge over the course of the experiment based on the morphological differences between the conditions. For the L1 Turkish participant group, I expected that positive transfer of L1 processing strategies, and potentially the transfer of transparent morphological structures from the L1, will benefit learning in the agglutinating condition. Therefore, I predicted learning rates to be higher in the agglutinating than in the fusional condition. This means that the learning curve in the agglutinating condition should be steeper, due to an earlier and steeper increase in accuracy in this condition, compared to the fusional one.

**Figure 3** shows mean accuracy across conditions by trial. As expected and observed previously, participants started at around chance level (25% correct due to random selection among 4 options) and generally improved over time. Towards the end of the experiment, participants reached relatively high accuracy of around 70% in both conditions. However, performance appears similar in both conditions across trials, which contradicts the prediction that the agglutinating system should be learned better. Mean accuracy across all trials for the agglutinating condition was 0.63

(SD=0.24), and slightly lower for the fusional condition (0.57, SD=0.22). Across the final 10 trials, mean accuracy for the agglutinating condition was 0.75 (SD=0.28), for the fusional condition 0.68 (SD=0.30). To test whether the rate of improvement differed between conditions I fit a logistic polynomial mixed-effects regression model as with the experiments reported in **Chapter 3**. The learning curves were modelled using second-order orthogonal polynomials, predicting correct answer by condition, trial, and their interaction. Trial was replaced by a linear and a quadratic time term; the time terms were added sequentially and the fit of the resulting two models (linear and polynomial) was compared using a Likelihood Ratio Test.

The inclusion of the quadratic time term significantly improved model fit ( $X^2=34.84$ ,  $p<0.001$ ). The polynomial model revealed a significant effect of the first time term ( $b=8.14$ ,  $SE=1.17$ ,  $p<0.001$ ), but not of the second time term ( $b=-0.80$ ,  $SE=0.55$ ,  $p=0.143$ ). This indicates that participants improved their accuracy over the course of the experiment and until the final trials. However, there was no significant effect of condition ( $b=0.18$ ,  $SE=0.21$ ,  $p=0.381$ ; this indicates the effect of condition when the two time terms are 0, which was at the midpoint of the experiment). Further, there was no significant interaction effect between condition and the first time term ( $b=0.16$ ,  $SE=1.11$ ,  $p=0.886$ ) or the second time term ( $b=-0.44$ ,  $SE=0.50$ ,  $p=0.371$ ). These results indicate that there was no difference in acquisition success across conditions over time, and both learning curves remained similar towards the end, where they did not flatten out significantly (model results are also shown in **Table 3** in the Appendix to this chapter). Overall, there is once again no indication of a learnability advantage of agglutinating over fusional morphology.

The L1 Turkish participant group was tested on a different recruiting platform than the L1 English participant group. However, the results of mean accuracy in both conditions reported here are in the range of those observed in the experiments in the previous chapter. It therefore appears that the mean learning rates themselves are not influenced by the platform used.



**Figure 3:** Mean accuracy by trial number and condition in the Experiment with Turkish native speakers. Error bars indicate 95% bootstrapped confidence intervals. Participants start at around chance level (0.25) on trial 1 and improve over time at a similar rate in both conditions. A comparison of the L1 Turkish to the L1 English group was performed together with the data from the L1 Spanish speakers, and is reported at the end of **Chapter 5**.

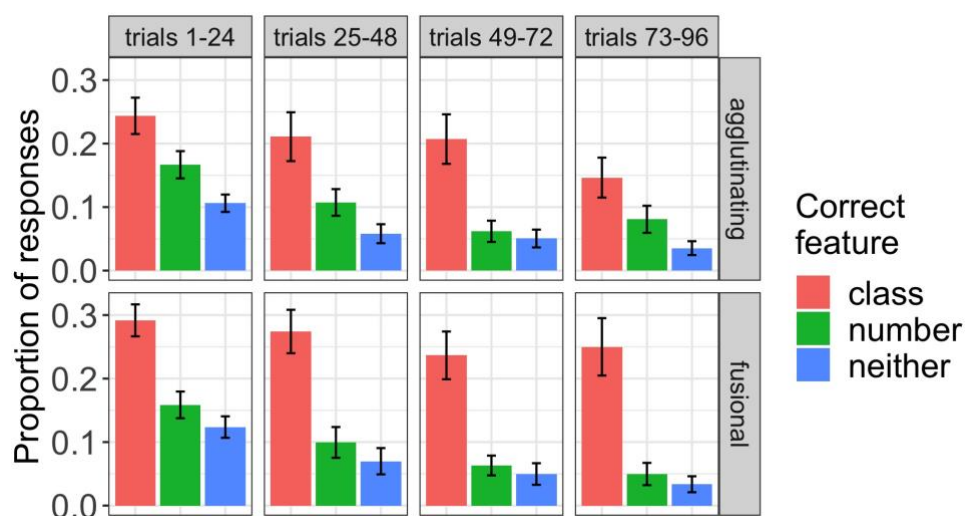
### 4.3.2 Discussion

The results reported above clearly contradict the prediction that positive transfer from the Turkish L1 will lead to acquisition ease in the agglutinating condition. However, transfer of morphological processing strategies has been shown to be beneficial in several studies (e.g., Havas et al., 2015; Portin et al., 2008). If transfer happened in this experiment, its effect was relatively small, since performance did not significantly diverge between conditions across trials. These results are unexpected, since L1 transfer is an additional predictor of learning ease in the agglutinating condition, beyond the transparency of its morphemes and the unavoidable inbuilt frequency difference of morphemes, which was in favour of the agglutinating over the fusional system. It remains to be seen whether Turkish native speakers nevertheless showed a slight transparency bias which can be detected in the analysis of error types in both conditions. This is reported on below.

### 4.3.3 Analysis of error patterns

As in the previous experiments, the four available choices in each trial corresponded to the four possible combinations of grammatical meanings, namely animate+singular, animate+plural, inanimate+singular and inanimate+plural. Therefore, it was again

possible to classify responses as either correct, CLASS correct only, NUMBER correct only, or none correct (e.g. animate+plural in a singular+inanimate trial). This applied to both conditions. Given that this time, participants' native language was highly transparent, it is possible that even though there was no overall learnability difference between conditions, their bias for transparent structures is stronger, manifesting as significantly more CLASS correct only responses in the fusional than in the agglutinating condition. For reasons discussed in **Chapter 3**, focusing solely on CLASS and enforcing a 4-way distinction within this feature would render the fusional morphemes transparent.

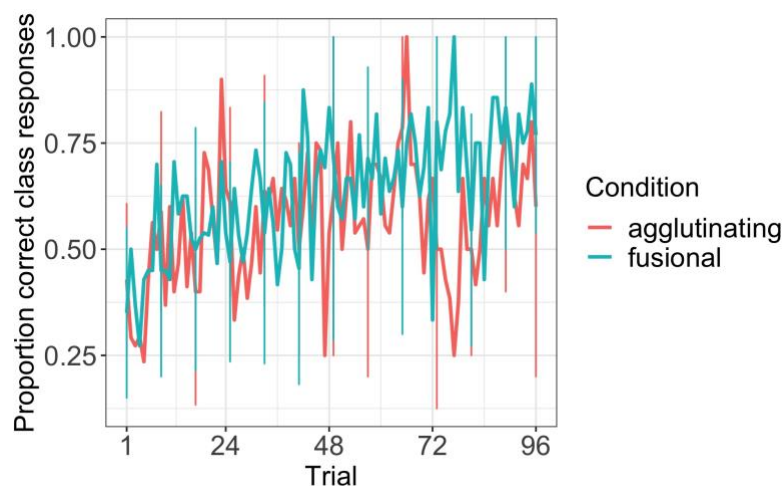


**Figure 4:** Proportion of the three possible types of error (CLASS correct only, NUMBER correct only, none correct) per condition, split up across four sections of the experiment with L1 Turkish participants. Note that only the subset of incorrect answers is shown.

Comparing the three types of errors, CLASS correct only responses were much more frequent than the other two types, and they seem to be even more frequent in the fusional than in the agglutinating condition. **Figure 4** illustrates this distribution, showing that CLASS correct only responses remain dominant throughout the experiment. For instance, participants would choose inanimate singulars in trials showing inanimate plural images, rather than choosing animate singulars. **Figure 5** displays the proportion of these CLASS correct only responses among incorrect answers across trials in both conditions. As in the analysis of error types reported in **Chapter 3**, I assessed the difference between conditions in CLASS correct only responses using a logistic polynomial mixed-effects regression model fit to the subset

of the data with incorrect answers. The dependent variable was CLASS correct. I again used second-order orthogonal polynomials with fixed effects of condition, trial number (replaced by a linear and a quadratic time term) and their interactions. Condition was effects-coded whereby agglutinating was compared to the grand average. Random by-item intercepts were removed from both models due to convergence failure.

The inclusion of the quadratic time term significantly improved model fit ( $X^2=16.89$ ,  $p=0.005$ ). The polynomial model revealed a significant effect of the first time term ( $b=3.97$ ,  $SE=0.97$ ,  $p<0.001$ ) and a marginal effect of the second time term ( $b=-1.27$ ,  $SE=0.65$ ,  $p=0.051$ ), indicating that among incorrect responses, CLASS correct only responses increased over the course of the experiment and that the curve of such responses somewhat flattened out towards the end. The effect of condition was also marginally significant ( $b=-0.38$ ,  $SE=0.21$ ,  $p=0.067$ ) and there was again a marginally significant interaction between condition and the first time term ( $b=-1.65$ ,  $SE=0.96$ ,  $p=0.087$ ). This indicates that CLASS correct only responses tended to occur more often in the fusional condition than in the agglutinating condition and that across trials, CLASS correct only responses reduced somewhat more rapidly in the agglutinating condition compared to the grand mean. The interaction between condition and the second time term was not significant ( $b=0.23$ ,  $SE=0.64$ ,  $p=0.721$ ), indicating that CLASS correct only responses slightly decreased towards the end of the experiment in a similar way across conditions. Full model results are shown in **Table 4** in the Appendix to this chapter.



**Figure 5:** Mean proportion of incorrect responses in which only CLASS was correct, by trial number and condition. Error bars indicate 95% bootstrapped confidence intervals. Only the proportion among incorrect answers is shown.

#### 4.3.4 Discussion

The assumption with regards to transparent morphology in the Turkish L1 was that it will lead to decomposition ease and potentially a bias for transparent structures in the L2, and therefore to significantly higher learning rates in the agglutinating compared to the fusional condition. However, overall accuracy rates in this experiment are very similar to those of the previous experiments with English native speakers: the artificial language in both conditions is learned equally well overall and there are no apparent differences between conditions at any point during the experiment. When investigating the types of errors made, the picture is again similar to that of the previous experiments in that participants show a stronger focus on CLASS than NUMBER in general, and this effect increases as the experiment progresses. There is also a slight tendency that the effect is stronger in the fusional than in the agglutinating condition. Note that this means that participants learned CLASS better than NUMBER, which is mirrored by participants' responses in the questionnaire. Many report ways of splitting CLASS into four different categories, such as large and small animals vs. tools and personal belongings (see **Section 4.3.5.5**), ignoring NUMBER altogether. This essentially turns the fusional morphemes into transparent morphemes, since they only express one feature, namely CLASS. It also explains why the tendency of CLASS correct only responses is stronger in the fusional condition, since agglutinating morphemes are already transparent in this respect. Further, the results from Experiments 1 and 2 in the previous chapter have indicated that participants are segmenting the two grammatical agglutinating morphemes well. The learning rates in Experiment 4 appear similar to these, which suggests that also Turkish native speakers segmented the morphemes (at least offline), and could theoretically benefit from the hypothesised advantage of transparency in the agglutinating system.

The focus on CLASS is only marginally more pronounced in the fusional compared to the agglutinating condition, which is not exactly what was expected – I assumed that native speakers of Turkish would show a similar learning pattern to native speakers of English, namely to produce significantly more CLASS correct only responses than any other error type. Given that Turkish displays ANIMACY distinctions to some degree in the area of DOM, the difference in the focus on CLASS between conditions was expected to be at least as big as for the L1 English group, due to potential transfer of the abstract category of ANIMACY (Luk & Shirai, 2009; Sabourin et

al., 2006). Another expectation was that pervasive L1 experience with morphemes that only express one grammatical meaning will lead participants to assume that morphemes in the fusional condition tested here are also transparent, and that this bias will be stronger in the Turkish native speaker group compared to the L1 English group. However, the difference between the agglutinating and the fusional condition regarding CLASS correct only responses is only marginally significant, and it seems as if the slight bias for transparent structures that was previously observed in the error type analyses was not amplified by the transparent L1 in the present experiment. This observation and its potential implications will be discussed further in the **General Discussion** in **Section 4.4**. Before this, the next section will outline some trends and recurrent insights from the questionnaire responses.

#### **4.3.5 Analysis of questionnaire**

Although the analysis of the experimental data obtained is largely quantitative, participants' responses in the short final questionnaire are revealing because they support many of the insights reported in the previous results section. For this reason, an overview of commonly reported strategies of learning the meaning of the grammatical morphemes will be provided. This includes examples of participants learning only one of the two grammatical features, participants learning the correct grammatical system, and the strategies of classifying the images in the trials. I will also summarise some other strategies which deviate from the dominant ones, but which re-occurred among participants as well as experiments more generally. The ways that participants classified the trial images are especially interesting because they highlight how some participants adjusted their strategies based on the feedback they received during trials and how often, participants paid more attention to CLASS over NUMBER. Overall, the percentage of participants providing responses in the text boxes of the questionnaire was larger among L1 Turkish than L1 English participants, and their answers were generally more elaborate. The question that participants responded to was the Turkish equivalent of "How did you decide which button to click? Did you

develop any strategies and can you describe the pattern of the language?” and the reported responses are the English translations of the Turkish original responses<sup>15</sup>.

#### **4.3.5.1 Participants focusing on suffixes rather than stems**

At first, it is important to note that many of the questionnaire responses suggest that participants learned to ignore word stems and only focused on the suffixes. This was intended by using 96 different stems, and the responses by participants shown below explicitly mention that such a strategy was used:

“In the first 4-5 questions I discovered that the suffixes were added depending on whether the object is singular or plural and animate or inanimate. I didn’t pay attention to the roots of the words.” (B16, fusional)

“I only followed two rules. The second bit of the words referring to animate beings started with m (like mupi) and with l for objects (like lupi). If singular things the word ends with i, if plural with a. Thus I didn’t look at the first part of the words at all after the first few questions.” (C21, agglutinating)

#### **4.3.5.2 Participants reporting the correct classification system**

Moving on to image classification strategies, some participants reported the correct classification between CLASS and NUMBER and how they mapped the distinction onto the grammatical morphemes. A few quotes serve as illustration of typical responses:

“I grouped them by the words’ last two letters. The ones ending with lo/mu are for animals and ones ending with pi/ka are for inanimate objects. Lo and pi singular, pi and ka plural.” (A6, fusional)

“The words ending with -ka are used for plural animate beings, the ones ending with -mu are used for plural inanimate objects, the ones ending with -pi are used for singular animate beings, the ones ending with -lo are used for singular animate beings. After a certain point in the test, I responded according to those rules.” (A12, fusional)

---

<sup>15</sup> Translations of the questionnaire responses from Turkish to English were provided by one of my Turkish informants, who also checked whether any of the written responses indicated that a participant was not a native speaker of Turkish. However, such a case did not occur.

#### **4.3.5.3 Difficulties assessing meaning-morpheme mappings employed by participants**

However, some participants' answers only mention the image classification system and not the grammatical morphemes, which makes it impossible to know how (and how consistently) they mapped the grammatical meanings onto morphemes. Further, in such cases in the agglutinating condition, it is not possible to infer whether participants treated the agglutinating morphemes in a segmented fashion or as a chunk. An example is provided below, which, however, suggests segmentation of at least the NUMBER suffixes, which followed the CLASS suffixes in the artificial words.

“I discovered a pattern about the objects being animals or objects and quantity of the shown things. This pattern affected the last two syllables of the words. I selected according to this after the first few questions.” (D15, agglutinating)

However, answers as the one above are still revealing as to whether participants distinguished animate and inanimate as well as singular and plural, or whether they used other classification systems. Other responses provide less insight, such as participants only commenting on the part of the word they focused on, as illustrated below. Here, it is unclear how objects were classified and what meanings were ascribed to the grammatical morphemes. Often, however, answers were more descriptive of the strategies used.

“I chose depending on the last syllables of the words.” (A10, fusional)

Finally, a few participants did not provide an answer to this particular question in the questionnaire. From those who did, a selection will be presented below that illustrates different pathways of learning during the experiment.

#### **4.3.5.4 Participants who only learn one of the two features**

Besides participants who learn the two features successfully, some only learned one of them by the end of the experiment. The first answer below shows how a participant only learned NUMBER, but not CLASS, while the second answer illustrates the more common pattern, namely learning CLASS, but never learning NUMBER.

**Example of a participant only learning NUMBER, but not CLASS:**

“-pika singular, -pilo plural, -muka singular, -mulo plural. I figured out these were the suffixes for plural and singular words.” (D9, agglutinating)

**Example of a participant only learning CLASS, but not NUMBER:**

“I classified them as animals and objects. Objects ended with -ka and -lo suffixes but couldn’t figure out the further distinction between these two. I thought about objects being classified as “tools” or “personal belongings”, but some counter-examples were present. In animals I tried to find patterns based on the species, size or their habitats but this seemed to not work as well (like birds ending with -pi at first but then with -mu around the end). Overall, I tried to look at the last syllable of words and use patterns between images and select an option accordingly.” (D7, fusional)

In addition, other participants eventually learned both features successfully, but they acquired one feature earlier than the other one. The first quote below illustrates how a participant first noticed NUMBER and later the CLASS-based distinction. The opposite pattern was much more common, namely participants noticing CLASS before NUMBER, as illustrated in the second quote. The third and fourth quotes show how some participants seemingly noticed both features simultaneously at some point in the experiment.

**Example of participant noticing NUMBER before CLASS:**

“It was easy to distinguish the categorisation by singular and plural. It took me a bit longer to realise the animate/inanimate distinction, especially on the inanimate bit.” (B34, fusional)

**Example of participant noticing CLASS before NUMBER:**

“Around the end I realised that animals and objects had different suffixes depending on whether they were singular or dual.” (D10, fusional)

**Example of participants noticing both features at the same time:**

“Yes, after the first 10 images I realised there were different suffixes for animate/inanimate and singular/dual categories and used this shortcut for the rest of the questions.” (E11, fusional)

“About halfway through the test I realised -mu = inanimate, -lo = animate, -pi = plural, -ka = singular.” (E9, agglutinating)

Note that while participants were still learning the system, they naturally generated errors of different kinds as illustrated in **Figure 4**. Since it was more common for

participants to acquire CLASS before NUMBER based on questionnaire responses, it links with the observation that CLASS correct only responses are also the most dominant type of error in **Figure 4**.

#### **4.3.5.5 Examples of different image classifications used**

Given this discrepancy between the two grammatical features and the fact that not all participants succeeded in learning the artificial language, it remains to be seen how they attempted to classify the images shown in the trials. In their search for patterns between the word endings and the content shown in the pictures, participants considered an impressive amount of different hypotheses. The selection of examples reflects the fact that virtually no participant attempted a classification within NUMBER beyond singular – plural (occasionally also referred to as dual). Within the domain of object categories, however, participants considered grouping items according to animate/inanimate and a secondary distinction within each of these. Examples are animal size, habitat and diet for animates as well as clothes, tools, size and personal belongings for inanimates. Some participants also report considering GENDER. Participant C28 did not attempt any secondary distinction and chose answers randomly between two options. The subsequent quotes are an illustration of the wide range of hypotheses tested, but they also show the general tendency that participants explore options within the CLASS feature much more than considering a second, different feature simultaneously. Additionally, participants who did not consider NUMBER as a feature neither considered any other hypothetical feature aside from CLASS.

“Wild or domestic animal, big or small animal, wearable or non-wearable things.” (B7, fusional)

“When distinguishing the animate or inanimate beings, I paid attention to the -mu and -pi suffixes in the second syllable. I figured out that -mu was used for inanimate beings and -pi was used for animate beings. But I could not figure out the distinction within the animate and inanimate beings. I tried to categorise the inanimate things by textile, clothing materials etc. but could not find an accurate pattern. For animals I tried to categorise them by their habitats: land, sea or flying animals but could not form an accurate strategy.” (B35, agglutinating)

“If the word ends with pi or lo it indicates an animal in the image, -pi if a heavy animal with four feet. -lo or -mu if it’s an inanimate object. If it’s an object used in daily life the word ends with -ka.” (C23, fusional)

“I tried to pay attention to see whether the words had a gender. I tried to categorise animals and objects. Also paid attention to the size of the objects and whether they were small or big.” (C24, fusional)

“I chose -pi or -ka randomly. I gave them both 50% chance, the same strategy for objects, -lu or -mo. [...]” (C28, fusional)

“Even though it seems like small animals get the -pika suffix, sometimes large and wild animals also had the -pika suffix, and I realised -pilo was used for birds. There might be a difference depending on the gender or the eating habits of animals too (herbivore/carnivore). For big birds -pilo wasn't used.” (E1, agglutinating)

“I realised that there were different suffixes used for animals and objects. But there were two different suffixes for both animals and objects. I tried to categorise objects as things we wear, things we use while eating etc. to find which suffix was used. Same with animals according to their habitats and their species but wasn't so successful.” (E22, fusional)

#### **4.3.5.6 Participants focusing on single letters**

A few additional strategies seem noteworthy, since they deviate from the dominant patterns. Some participants reported to have paid attention to single letters instead of syllables. While this would have allowed them to choose answers correctly, it suggests that not every participant perceived the grammatical morphemes as units in the input. Three example responses are shown below.

“last letter a: inanimate plural, last letter i: inanimate singular, last letter o: animate plural, last letter u: animate singular.” (A19, fusional)

“I thought the last letters A-U indicated the plural-singular distinction, and the syllable before it starting with L-P indicated the animate-inanimate distinction and I tried to select the options according to that.” (B10, agglutinating)

“u -> singular, animal; a -> plural, animal; o -> singular, object, i -> plural, object.” (C5, fusional)

#### **4.3.5.7 Participants' awareness of the agglutinating nature of the artificial language**

As a final remark, some participants commented on the fact that the language they were learning was agglutinating, demonstrating an awareness of the linguistic structures in question. While the number of participants making such remarks was too small for an additional analysis, it might be insightful for future studies to see whether

morphological awareness can predict learning ease of different morphological structures.

“In the first 4-5 questions, I clicked on a random option to figure out the roots and suffixes. Afterwards realised that there were different suffixes for animate/inanimate objects and singular/plural. It appeared to be an agglutinative language. The first two syllables for the root of the word and the following -mu/-lo indicated inanimate-animate distinction respectively. Whereas -pi/-ka indicated singular-plural distinction respectively. I made some mistakes when I was distracted.” (D6, agglutinating)

“I realised I was dealing with an agglutinative language. Depending on the plural/singular and animate/inanimate state of the objects shown, the words took suffixes like -lopi or -muka and answered the rest of the questions according to this strategy.” (E16, agglutinating)

In sum, this overview of different types of questionnaire responses illustrates how participants strongly focus on the grammatical morphemes and partly report them in a segmented fashion in the agglutinating condition. It was also shown how some participants notice the classification system into two binary grammatical features earlier than others. Depending on how many trials participants need to learn this system, they naturally produce more or fewer errors in the process. Further, it became apparent how CLASS was learned more easily than NUMBER and how participants tried different hypotheses to turn CLASS into a category with four values. Finally, a few participants showed metalinguistic awareness and noticed that the artificial language was agglutinative in nature. Overall, these findings confirm that the intended experimental design worked well and they reflect the strong focus on CLASS as it is revealed in the error type analysis.

#### **4.4 General Discussion**

The aim of this chapter was to complement the picture of results of **Chapter 3** with regards to the role of experience with morphological decomposition (leading to processing transfer onto the L2) and potentially a bias for transparent morphemes when acquiring L2 morphology. It was predicted that the L1 Turkish group would also learn CLASS better than NUMBER, but that the proportion of CLASS correct only responses may be even larger than for native speakers of English, due to a stronger bias for transparent structures. Native speakers of Turkish were also expected to perform

better overall due to their larger morphological awareness, resulting in higher overall learning rates in both conditions, compared to native speakers of English. Lastly, the L1 Turkish group was predicted to learn significantly better in the agglutinating condition.

These predictions were only partly confirmed by the data analysis. At first, the main prediction of better learning in the agglutinating condition is not confirmed by the data, since the effect of condition on overall accuracy is not significant. Further, while CLASS was learned better than NUMBER as predicted, as is also reflected in the questionnaire responses, CLASS correct only responses were only marginally more frequent in the fusional compared to the agglutinating condition. The overall accuracy rates also appear to be in the range of those for L1 English speakers, which again contradicts the predictions. This will be followed up with another statistical model in **Section 5.4**. Overall, what these results suggest is that the morphological make-up of the Turkish language did not exercise any effects on learning in the two conditions that appear different from those in the experiments with L1 English speakers. This is because there appears to be neither an overall bias for agglutinating structures (which would be reflected in higher overall accuracy in the agglutinating condition) nor a slight bias visible in the error type analysis which is stronger than with previous participants. The latter would have meant that Turkish-speaking participants assume more strongly that the fusional morphemes are transparent and that they explored CLASS much more than NUMBER. Since the overall learning rates do not seem to differ much from Experiment 1 either, this suggests that if Turkish-speaking participants used morphological decomposition to a larger extent either online or offline compared to the L1 English group, it did not affect learning in the experiment.

Research on morphological processing at the onset of learning a second language is sparse, and it is possible that in the present experiment, the Turkish L1 did not exercise any or only a small learning advantage. Previous studies did focus on learner groups that were more advanced than complete beginners (Portin et al., 2007; Portin et al., 2008; Gor & Jackson, 2013). However, Havas et al. (2015) also tested learning of an artificial language and observed an effect of experience with morphological decomposition. Thus, it seems possible that L1 decomposition strategies help with the acquisition of L2 morphology under certain circumstances, but not in this particular experiment. It also appears as if the agglutinating morphology of Turkish did not lead to transfer of this particular structure in the sense of an abstract

category (Sabourin et al., 2006) or to a bias for transparency of any other sort (Igartua, 2015).

However, also a different type of explanation needs to be considered. The artificial language experiment was not designed to test acquisition of agglutinating morphology, as is found in Turkish, but one aspect of agglutination, namely morpheme transparency. If transparency is a predictor of learning ease, then the Turkish-speaking participants of this experiment would have certainly been expected to reflect this in their learning pattern – their L1 is highly transparent and lines up with the language in the agglutinating condition. Further, some bias is visible in their focus on CLASS in the fusional condition, and there might have been some processing facilitation. There is no evidence either that the choice of grammatical features interfered with learning for this participant group. Lastly, the colour manipulation in Experiment 2 tested whether offline segmentation was an issue during Experiment 1 with native speakers of English, but accuracy rates were not altered by it, suggesting that segmentation happens even without additional cues. Since the learning rates in Experiment 4 appear similar to those of Experiment 1, we can assume that segmentation of the two agglutinating morphemes was again successful. However, it clearly did not lead to better learning in this condition. For these reasons, it appears to be the case that transparency is not a main factor that makes morphology more learnable, since not even the participant population who should benefit most from transparency does so. Augmented with the insights from **Chapter 3** where the proposed learning advantage was missing as well, the present results offer accumulated evidence against the suggested impact of transparency (Bittner et al., 2003; Brown, 1976; Goldschneider & DeKeyser, 2001; Hengeveld & Leufkens, 2018; Igartua, 2015).

It needs to be remembered that the L1 Turkish participant group was a bit smaller and tested on a different recruiting platform than the L1 English participant group, which could have caused differences in learning behaviour. However, the results of mean accuracy in both conditions reported here are in the range of those observed in the experiments in **Chapter 3**. It therefore appears that the mean learning rates themselves are not influenced by the platform used. Nevertheless, **Chapter 5** will present the testing of a third group of participants, namely native speakers of Spanish, again via Amazon Mechanical Turk as with the initial experiments. If experience with inflectional morphology is an important factor in L2 learning, then

these results should reveal a difference to the learning rates from native speakers of English, who were tested via the same platform.

In short, there were several reasons to expect a learnability advantage of the agglutinating system due to the participants being native speakers of Turkish, an inflectionally rich agglutinating language. Beyond the transparency of the morphemes in the agglutinating condition, participants' experience with agglutinating morphology increased the expectation of better learning. This could have supported learning in various ways, either through a transfer of processing strategies, which should help with morpheme identification and learning of the grammatical categories, or through familiarity with this type of morphological structure. Some participants even reported that they were learning an agglutinating language, which demonstrates meta-awareness. It might also be the case that there is a more general preference for the type of morphological structure prevalent in the L1. In this experiment, this overlapped with the claim that transparent structures are in general easier to learn. Finally, there was also an unavoidable difference in the frequency of agglutinating compared to fusional morphemes, favouring the agglutinating condition. However, the slight bias for transparent structures that we see in the error type analysis does not seem to be amplified by the transparent L1.

Since the transparency difference between the experimental conditions was very clear and the main difference between them, why do we not observe facilitation in the agglutinating condition for Turkish native speakers? One possibility is that Turkish speakers are not sensitive to the transparency of Turkish morphology, which, however, appears less plausible. Instead, it is conceivable that speakers are sensitive to transparent morphemes, but there is no strong bias for them or that in this paradigm, transparent structures are not advantageous over non-transparent ones. In sum, the lack of better learning in the agglutinating condition points to the fact that transparency is not providing a learning benefit.

Another expectation was that pervasive L1 experience with morphemes that only express one grammatical meaning will lead participants to assume that morphemes in the fusional condition tested here are also transparent, and that this bias will be stronger in the Turkish native speaker group compared to the L1 English group. However, the difference between the agglutinating and the fusional condition regarding CLASS correct only responses in Experiment 4 is only marginally significant, and it seems as if the slight bias for transparent structures that was previously

observed in the error type analyses was not amplified by the transparent L1 in the present experiment. This observation and its potential implications will be discussed further in the **General Discussion** in **Chapter 6**.

## **4.5 Conclusion**

The analysis of the present chapter contributed a second group of L1 participants that allows us to see whether experience with morphological structures from the L1 impacts the acquisition of L2 morphology. This is in line with DeKeyser's (2005) description of a research path that tries to illuminate the interplay of L1 and L2 characteristics. Here, the research question is whether we can find an advantage of morphological transparency in a controlled artificial language learning experiment where we exclude additional factors. The nature of the Turkish language contributed additional reasons why the agglutinating system should have been easier to learn, since it is inflectionally rich and highly transparent. However, the results from native speakers of Turkish are very similar to those from native speakers of English in that no advantage of transparent structures was found and a bias for transparency appears weak at best. A third group of participants will be tested in the next chapter, but the evidence collected so far strongly points to the fact that transparency is not a strong predictor of acquisition ease.

## 4.6 Appendix

### Full model results:

	Estimate	Std. Error	z value	Pr(>  z  )
(Intercept)	1.12	0.16	7.13	<0.001
condition1	0.11	0.16	0.68	0.494
ot1	7.73	0.70	11.11	<0.001
ot2	-2.22	0.34	-6.51	<0.001
colorcoding1	-0.29	0.16	-1.85	0.064
L21	-0.07	0.15	-0.45	0.653
condition1:ot1	0.71	0.67	1.06	0.290
condition1:ot2	-0.12	0.32	-0.38	0.704
condition1:colorcoding1	0.07	0.16	0.46	0.647
ot1:colorcoding1	-0.94	0.67	-1.40	0.162
ot2:colorcoding1	0.52	0.32	1.64	0.101
condition1:L21	0.24	0.15	1.53	0.127
ot1:L21	-0.45	0.67	-0.67	0.500
ot2:L21	-0.18	0.32	-0.57	0.566
colorcoding1:L21	-0.11	0.15	-0.73	0.467
condition1:ot1:colorcoding1	0.42	0.67	0.62	0.536
condition1:ot2:colorcoding1	0.29	0.32	0.90	0.368
condition1:ot1:L21	0.87	0.67	1.31	0.191
condition1:ot2:L21	-0.15	0.32	-0.48	0.631
condition1:colorcoding1:L21	0.15	0.15	0.96	0.337
ot1:colorcoding1:L21	0.12	0.67	0.18	0.858
ot2:colorcoding1:L21	0.28	0.32	0.88	0.381
condition1:ot1:colorcoding1:L21	0.30	0.67	0.45	0.654
condition1:ot2:colorcoding1:L21	0.12	0.32	0.37	0.708

**Table 2:** Results from the combined data of Experiments 1 and 2 (English native speakers), overall accuracy. The predictors of condition, L2 experience and colour-coding were effects coded with agglutinating, some L2 experience and colour being compared to the grand average.

	Estimate	Std. Error	z value	Pr(>   z  )
(Intercept)	0.96	0.21	4.49	<0.001
condition1	0.18	0.21	0.88	0.381
ot1	8.14	1.17	6.94	<0.001
ot2	-0.80	0.55	-1.46	0.143
condition1:ot1	0.16	1.11	0.14	0.886
condition1:ot2	-0.44	0.50	-0.89	0.371

**Table 3:** Results from Experiment 4 with Turkish native speakers, overall accuracy. The predictor of condition was effects-coded with agglutinating being compared to the grand average.

	Estimate	Std. Error	z value	Pr(>   z  )
(Intercept)	0.59	0.21	2.85	0.004
condition1	-0.38	0.21	-1.83	0.067
ot1	3.97	0.97	4.08	<0.001
ot2	-1.27	0.65	-1.95	0.051
condition1:ot1	-1.65	0.96	-1.71	0.087
condition1:ot2	0.23	0.64	0.36	0.721

**Table 4:** Results from Experiment 4 with Turkish native speakers, CLASS accuracy in the subset of the data with incorrect responses. The predictor of condition was effects-coded with agglutinating being compared to the grand average.

## Chapter 5

### Testing the effect of transparency in artificial language learning in native speakers of Spanish

#### 5.1 Introduction

In **Chapter 4**, I described the need to test native speakers of different languages, which will allow us to observe the potential role of the morphology in the L1 for the acquisition of second languages. Studies have shown that transfer from the L1 onto the L2 occurs in various respects, including transfer of abstract categories and transfer of processing strategies (Portin et al., 2007; Portin et al., 2008; Sabourin et al., 2006). Within the scope of this thesis, it is most important to understand whether experience with either transparent or non-transparent morphology in the L1 will have an impact on how participants learn (non-)transparent morphology in the L2. For this reason, the artificial language learning experiment was conducted with native speakers of English (an inflectionally poor language) and with native speakers of Turkish (an inflectionally rich and morphologically highly transparent language). What remains to be tested is a group of participants who are native speakers of an inflectionally rich, but non-transparent language. In this chapter, this picture will be complemented by a participant group of native Spanish speakers. Spanish is inflectionally much richer than English and often displays fusional elements, therefore allowing me to test whether knowledge of fusional morphology from the L1 impacts learning in my experiment. Moreover, Spanish has partly semantic and partly grammatical GENDER and overtly marks an ANIMACY distinction in Differential Object Marking (DOM), suggesting that native speakers of Spanish may be more familiar with ANIMACY than native speakers of English or Turkish. While this is a coincidence of the L1 selected, the grammatical knowledge of the Spanish speakers will make it possible to test whether a) experience with fusional morphology leads to improved performance in the fusional condition or b) whether Spanish native speakers explore ANIMACY even more than native speakers of English or Turkish. This can provide evidence for the kind of linguistic knowledge that can be transferred onto an L2. In the section below, I will first describe the relevant aspects of Spanish morphology and grammar and then outline the predictions I derive for L1 Spanish speakers' performance in the experiment.

## 5.1.1 Relevant aspects of Spanish morphology and grammar

### 5.1.1.1 Spanish as a synthetic fusional language

Similar to Turkish, Spanish is a synthetic language and rich in inflections (Fábregas, 2018). Grammatical categories which are marked by inflectional means include TENSE, ASPECT, MOOD, PERSON and NUMBER on verbs, GENDER and NUMBER on nouns, adjectives, determiners and quantifiers, and superlative inflection on adjectives. According to Moreno-Sandoval & Goñi-Menoyo (2002), the Spanish verb has over 50 synthetically inflected forms. In addition to being synthetic, Spanish also displays many elements of fusional morphology as discussed in **Section 1.2**, including cumulative exponence, extended exponence, syncretism, inflectional classes and allomorphy (Fábregas, 2018; Moreno-Sandoval & Goñi-Menoyo, 2002). Cumulative exponence of PERSON and NUMBER is illustrated in the verbal paradigms in (1) and (2). Here, the suffixes in both tenses all jointly express both features, but this fusional characteristic can be found more generally across the verbal inflectional system in Spanish (Moreno-Sandoval & Goñi-Menoyo, 2002). An example of syncretism in Spanish is given in (3), where the suffix *-ba* can express both first and third person, a pattern of syncretism which applies to Spanish imperfective past forms in general (Fábregas, 2018).

(1) The verb *amar* ('love') in Spanish (present indicative)

1 <sup>st</sup> SG	<i>am-o</i>
2 <sup>nd</sup> SG	<i>am-as</i>
3 <sup>rd</sup> SG	<i>am-a</i>
1 <sup>st</sup> PL	<i>am-amos</i>
2 <sup>nd</sup> PL	<i>am-áis</i>
3 <sup>rd</sup> PL	<i>am-an</i>

*(adapted from Moreno-Sandoval & Goñi-Menoyo, 2002)*

(2) The verb *amar* ('love') in Spanish (future indicative)

1 <sup>st</sup> SG	<i>am-aré</i>
2 <sup>nd</sup> SG	<i>am-arás</i>
3 <sup>rd</sup> SG	<i>am-ará</i>
1 <sup>st</sup> PL	<i>am-aremos</i>
2 <sup>nd</sup> PL	<i>am-aréis</i>
3 <sup>rd</sup> PL	<i>am-arán</i>

*(adapted from Moreno-Sandoval & Goñi-Menoyo, 2002)*

(3) *cant-a-ba*  
sang.IMPV.1/3SG  
I/he/she sang

(adapted from Fábregas, 2018)

### 5.1.1.2 GENDER and ANIMACY distinctions in Spanish (and Basque)

Apart from being fusional, Spanish exhibits two types of CLASS distinction, GENDER and ANIMACY. This was not the motivation for selecting Spanish as the third L1 to be tested in this series of experiments, but participants' knowledge of these features needs to be accounted for when making predictions about their acquisition patterns. Both will be illustrated briefly, before deriving predictions as to what learning patterns the particular type of linguistic knowledge among L1 Spanish participants may produce in the experiment. First, Spanish is well-known for marking GENDER, although it is partly disputed whether stem-final vowels strictly indicate GENDER or an inflectional class (Mel'čuk, 2013; Fábregas, 2018). Spanish distinguishes between the feminine and the masculine gender; the ending *-a* on nouns typically indicates feminine and the ending *-o* typically indicates masculine. Examples are *cas-a* (house) and *jueg-o* (game), although there are exceptions. Nouns that end in *-e* or in a consonant are less predictive of the GENDER they take. Most nouns are only assigned one GENDER, but some nouns can take both endings, chiefly those denoting humans and animals, such as *hij-o* (son) – *hij-a* (daughter), *león* (lion) – *leona* (lioness) (Fábregas, 2018; Havas et al., 2015). In these instances, the lexemes can inflect in four different ways to express female+singular, female+plural, male+singular, male+plural (Havas et al., 2015).

As these examples show, GENDER in Spanish is partly assigned based on the semantics of the noun, partly based on the phonology of the nominal endings and partly arbitrarily. Therefore, Spanish can be classified as a language in which GENDER is assigned based on both semantic and formal (arbitrary) criteria (Brooks et al., 1993; Corbett, 2013b). Independent of whether stem-final vowels as in *hij-a* are regarded as morphemes or as part of the stem, they indicate a binary GENDER feature in Spanish, which triggers agreement in adjectives, determiners, demonstratives and pronouns (Havas et al., 2015). One can therefore assume that native speakers of Spanish are familiar with the grammatical feature of GENDER and they may transfer this knowledge onto the language learning task of the experiment at hand. At least some facilitation appears possible, although the ANIMACY distinction in the experimental language was

purely semantically conditioned. Participants may try to find a correspondence between the suffixes and GENDER distinctions underlying the images. Further, GENDER is one example of a NOUN CLASS distinction which is inherent to the entities shown in the picture (whereas NUMBER is not inherent, since an entity can occur both in the singular and in the plural in the stimuli used in this thesis).

Another example of NOUN CLASS is ANIMACY, which is to some extent also marked explicitly in Spanish. More specifically, the marking of direct objects with the personal preposition *a* is semantically conditioned, largely by the ANIMACY and SPECIFICITY of the object in question (Chamorro et al., 2016). The use of a dative preposition with direct objects is also referred to as Differential Object Marking (DOM). When direct objects are both animate and specific, as in (4a), they require the preposition *a*. When the preposition is not supplied, the sentence is ungrammatical as in (4b). In other instances, such as with inanimate direct objects (independent of SPECIFICITY), supplying the preposition *a* renders the sentence ungrammatical, as shown in (5). However, Garcia Garcia (2018) reports that DOM may be more gradient than this. DEFINITENESS and SPECIFICITY of human objects play a role for whether prepositional *a* is used, but further distinctions can be made when the object is non-human. When objects are animate, but non-human, DOM is optional, but it is largely ungrammatical when the object is inanimate (Garcia Garcia, 2018).

(4a.) *María vio al niño esta mañana.*

*María saw to+the kid this morning*

b. \**María vio el niño esta mañana.*

*María saw the kid this morning*

“*María saw the kid this morning.*”

*(adapted from Chamorro et al., 2016)*

(5a.) *María vio una película/la película esta mañana.*

*María watched a movie/the movie this morning*

b. \* *María vio a una película/la película esta mañana.*

*María watched to a movie/the movie this morning*

“*María watched a movie/the movie this morning.*”

*(adapted from Chamorro et al., 2016)*

Further factors affecting the use of DOM have been investigated for Spanish, but they do not appear to have a bearing on performance in the present experiments. What is important to observe is that native speakers of Spanish must have a sensitivity for the placement of direct objects on an ANIMACY scale in order to be able to use prepositional *a* appropriately in Spanish. In turn, this might mean that L1 Spanish-speaking participants will explore possible categorisations of the stimuli images beyond animate-inanimate, if they also have difficulties learning the second feature of NUMBER. Thus, while Spanish does not overtly mark ANIMACY as a grammatical feature as in the stimuli used in this thesis, the concept of ANIMACY is important for speakers to be able to produce grammatically correct sentences.

In addition, some of the participants tested in this experiment were native speakers of both Spanish and Basque (further details are provided in **Section 5.2.1**). Basque also expresses the notion of ANIMACY in a restricted part of the grammar, namely in the formation of local cases of noun phrases (locative, ablative, allative as well as directional, terminative and destinative compound cases) (Hualde & Ortiz de Urbina, 2003). The distinction that is made is between animate and inanimate noun phrases. Animate comprises humans and large animals and inanimate comprises all others. Only the locative case inflections differentiate animate from inanimate noun phrases, leading to different morphological markers, with some dialectal variation (Hualde & Ortiz de Urbina, 2003). Aside from this, Basque has no noun classes and no grammatical GENDER, although it marks sex on some lexemes (Hualde & Ortiz de Urbina, 2003). Overall, ANIMACY effects on the grammar appear more restricted in Basque than in Spanish. However, in this experiment, participants who were native speakers of Basque were also native speakers of Spanish and therefore had cumulated experience with different types of NOUN CLASS as exhibited by their native languages. Thus, together with the knowledge of partly semantic, partly grammatical GENDER, the awareness of ANIMACY means that L1 speakers of Spanish (and Basque) are more experienced with NOUN CLASS distinctions than L1 speakers of English, and likely also more than L1 speakers of Turkish, where DOM is less prominent (and there is no GENDER marking). The possible implications of these grammatical features in the Spanish L1, along with Spanish being a relatively fusional language, will be discussed in the next section.

### 5.1.2 Predictions about learning outcomes for the L1 Spanish group considering language transfer

Based on the review of studies considering transfer and other facilitation effects from the morphology of the L1 onto the L2 in **Section 4.1.2** in the previous chapter, predictions can also be derived for the L1 Spanish-speaking group. Since the experiment with L1 Spanish speakers (Experiment 5) uses the same methodology as Experiments 1 and 4, as I will show in **Section 5.2**, all three L1 groups can be compared. This will be reported on in **Section 5.4**. Like Turkish, Spanish is a synthetic language and inflectionally richer than English. Relatedly, research has found that experience with morphological decomposition can lead to increased decomposition in an L2 compared to learners of a morphologically poorer L1 (Portin et al., 2007; Portin et al., 2008). A study by Havas et al. (2015) has found that native speakers of an inflectionally richer language, Finnish, show greater awareness of a morphological rule and better performance in a generalisation task with an artificial language than speakers of a language with fewer inflections, Spanish. In line with these findings, I assume that greater experience with morphology through the L1 leads to better identification of morphological patterns and retention of morphemic forms. If this facilitation effect also applies to early stages of learning as they are tested here, then native speakers of Spanish should perform better in both conditions than native speakers of English, but potentially not better than native speakers of Turkish (first prediction). Note that while this experiment does not test online processing, processing ease should nevertheless be reflected in higher accuracy rates. Of course, the predictions for L1 effects for Turkish were not borne out.

Next, if the morphological technique dominant in the L1 undergoes transfer effects like abstract categories, then we would expect Spanish speakers to learn better in the fusional condition. However, this prediction is less supported in the literature and it stands in opposition to the popular view that transparent structures are learned better than non-transparent structures (Aksu-Koç & Slobin, 1985; Bittner et al., 2003; Brown, 1976; Don, 2017; Hengeveld & Leufkens, 2018; Igartua, 2015; Penke, 2012). This is the strongest prediction guiding this experiment. For this reason, the effect of transparency is expected to be smaller in this group than in the Turkish-speaking group, but given claims about the benefits of transparency or a preference for agglutinating structures by L2 learners as claimed by Igartua (2015), we should still

see an advantage of the agglutinating condition. In sum, the individual predictions mean that the experience with synthetic fusional morphology should again allow participants to better exploit the benefits of the agglutinating condition, somewhat attenuated by their experience with the opposite exponence type (second prediction). Note again, however, that there was no evidence for a transparency advantage of the agglutinating condition in Experiment 4, where it was most expected.

Third, regarding transfer of abstract categories, I illustrated how ANIMACY and GENDER both exist in Spanish, and to a lesser extent in Basque. Although ANIMACY is strictly semantically conditioned in the artificial language, while it is partly formally conditioned in Spanish, some transfer of the abstract category is expected. Therefore, participants are expected to learn the ANIMACY distinction quickly, which should result in a higher number of CLASS correct only responses compared to the other two participant groups, at least in the earlier phase of the experiment. According to Sabourin et al. (2006), this is a case of deep transfer, since ANIMACY and GENDER marking differ between Spanish and the artificial language. Considering the claim made by Luk & Shirai (2009) that especially structural and semantic similarity of morphemes between L1 and L2 lead to acquisition ease, we would expect native Spanish speakers to learn NUMBER to a similar extent as native speakers of English and Turkish, since Spanish marks NUMBER similarly to both languages (plural marking through suffixation). Given that NUMBER is structurally and semantically more similar in the artificial L2 and in all three L1s compared to ANIMACY, findings by Luk & Shirai (2009) and Sabourin et al. (2006) would additionally predict that NUMBER is learned better than ANIMACY (surface vs. deep transfer). However, we have observed that English and Turkish native speakers focus much more on CLASS, which appears to be driven not by transfer, but by a bias for transparent morphemes. Thus, I again expect a stronger focus on CLASS than NUMBER, but the difference to be even stronger between both in this experiment with L1 Spanish speakers (third prediction). This prediction conflicts with the consideration that native speakers of Spanish might have a bias for fusional structures (in which case both grammatical features should be learned well), but note that better performance in the fusional condition is treated as less plausible based on the literature.

## 5.2 Methods

The experiment conducted with native speakers of Spanish (Experiment 5) is overall the same as Experiment 1 in **Chapter 3**. However, the stimuli and instructions were fully adapted to the Spanish language to ensure similar testing conditions.

### 5.2.1 Participants

Participants were recruited via Amazon Mechanical Turk. The experiment reported in this chapter was conducted prior to the one with Turkish native speakers for practical reasons. The numbers of Mechanical Turk workers situated in Spanish-speaking countries (for instance Spain and countries in South America) appeared large enough to warrant using Amazon Mechanical Turk as a platform again. It was also relatively easy to find Spanish native speaking colleagues who were able to help with the creation of stimuli and assist with text translations. The plan of testing three different participant groups, representing a predominantly isolating, agglutinating and fusional L1, respectively, was already in place during the conduct of the experiments reported in **Chapter 3**. The different order of reporting these experiments in this thesis was chosen for better illustration of the theoretical underpinnings of the L1-related hypotheses.

Since the number of available Amazon Mechanical Turk workers located in Spain was lower than those in the United States of America, the experiment was advertised to workers in Spain, Mexico, Chile, Argentina and Peru simultaneously. This was done to speed up data collection - since Amazon Mechanical Turk only allows to recruit participants based on their location, not their native language(s), it was easier to open the study to participants in various countries at the same time.

The use of Amazon Mechanical Turk to recruit native speakers of Spanish means that a difference in performance between native Turkish and native Spanish speakers could in part be due to the different recruiting platforms. In this case, a comparison of acquisition success between these two groups should be done cautiously. However, it will be possible to compare Spanish L1 speakers' learning rate in the agglutinating condition with that in the fusional condition. Additionally, any differences between the L1 English and the L1 Spanish groups cannot be due to the testing platform, since it was the same for both.

100 participants were recruited online over the course of 8 weeks via Amazon Mechanical Turk. They all self-reported as native speakers of Spanish and to be at least 18 years old. The 100 eligible participants were distributed across countries in the following way: Spain: 82, Mexico: 11, Chile: 2, Argentina: 4, Peru: 1. They were each paid \$4 for their time. Participants were randomly assigned to one of the two conditions (44 to the agglutinating condition and 56 to the fusional condition)<sup>16</sup>. I used the Mechanical Turk qualification and ID systems to ensure that participants were only able to participate in the experiment once. Note that despite the relatively high number of participants taking notes, it was possible to recruit 100 participants overall, which is similar to the number used in the experiments with English native speakers.

## **5.2.2 Creation of stimuli and additional experimental material**

### **5.2.2.1 Artificial stems**

As in all the previous experiments, the artificial language consisted of 96 stems and four suffixes, encoding the binary features NOUN CLASS (animate or inanimate) and NUMBER (singular or plural). The artificial stems were also adapted to the new participant group of native speakers of Spanish so that participants did not need to spend resources on learning novel phonotactics of the vocabulary. Together with a native Spanish-speaking colleague, I created pseudo-Spanish stems whereby we modified the original pseudo-English stems in minimal ways to arrive at stimuli that

---

<sup>16</sup> Of the 100 participants, 37 were bilingual according to the criterion used before, i.e. speaking more than one language from birth or from a young age, and being fluent in it. These participants were included in the analysis since the proportion was much larger than in the L1 English and L1 Turkish experiments. However, the bilingual participants were included as a separate group in the analysis. The native languages most often indicated aside from Spanish were Catalan, Basque, Galician – i.e. languages spoken in Spain, but also English. Languages indicated less frequently were French, Russian, Romanian and Ukrainian. 68 additional participants took notes during the experiment, based on their answers in the questionnaire following the experiment; their data was excluded. Another 19 participants were excluded for different reasons, such as selecting an incorrect answer in one of the native speaker tests, being classified as a non-native speaker of Spanish based on their questionnaire responses by one of the Spanish informants or re-starting the experiment after some trials had already been completed. The remaining data from 100 participants was included in the analysis. Of these, 92 participants indicated having at least basic skills in at least one other language apart from their native language(s). Note that Spanish-speaking participants have also been found to largely speak additional languages in a study by Havas et al. (2015).

followed Spanish phonotactics.<sup>17</sup> The pseudo-Spanish stems were also monosyllabic and did not exist as words in Spanish and the four morphemes used were *-mu*, *-ca*, *-pi* and *-lo* (the change from *-ka* to *-ca* was done to represent the Spanish alphabet). We ensured that the attachment of the grammatical morphemes did not yield any existing Spanish words, but a heavier restriction with the agglutinating morphemes was necessary. Since *mulo* and *loca* are existing words in Spanish, any randomisation of the assignment of meanings to the grammatical morphemes which would include these two suffix strings in the agglutinating condition had to be excluded. This reduced the number of possible randomisation mappings considerably, from 24 to 4, and the only agglutinating endings used were *-camu*, *-capi*, *-lomu*, *-lopi*. Despite this, the option with the reduced amount of randomisation was chosen to maintain consistency with the stimuli used in the previous experiments.

We again ensured that stems were not too similar to each other and that a variety of letters from the Spanish alphabet were used. Examples of pseudo-Spanish stems are: *breu*, *cos*, *mer*, *pois*, and *vri*. Finally, to mimic stress patterns in Spanish words, the trisyllabic agglutinating words (such as *jórlopi*) were accented on the first syllable, i.e. the stem, while the bisyllabic fusional words (such as *cetmu*) did not receive any written accents. The same stress patterns were reflected in the audio stimuli, which were also updated to fit the written stimuli and to follow Spanish pronunciation. The audio stimuli were created using the Mac speech synthesizer with voice Jorge.

The mappings of grammatical meanings (animate, inanimate, singular, plural) onto morphemes was done in the same way as in the previous experiments and transparency was manipulated across conditions as shown in **Table 1**. The exact mappings between morphemes and grammatical meanings were again randomised across participants in each condition, but followed the restrictions described above regarding the agglutinating condition.

---

<sup>17</sup> I would like to thank the four native speakers of Spanish, colleagues of mine, who were involved in the creation of the Spanish experimental material: Diana López-Lugo, Esperanza Ramos Badaya, José Segovia Martín and Mora Maldonado.

	animate singular	animate plural	inanimate singular	inanimate plural
agglutinating	grá-lo-mu	grá-lo-pi	grá-ca-mu	grá-ca-pi
fusional	gra-pi	gra-lo	gra-mu	gra-ca

**Table 1.** Illustration of the mapping of grammatical meanings onto suffixes in both conditions. *grá-* is the noun stem in each case (the accent is only used in the agglutinating condition due to words being tri-syllabic), hyphens are provided for illustration only and did not appear in the stimuli.

### 5.2.2.2 Translations and creation of additional material

All information and instructions used during the experiment were again translated with the help of native speakers so that participants were only working with Spanish and the artificial language. This includes the study information form which participants needed to accept in order to be able to participate as well as all instructions that participants received during the experiment and the final questionnaire. The study advert visible on Amazon Mechanical Turk was also exclusively kept in Spanish.

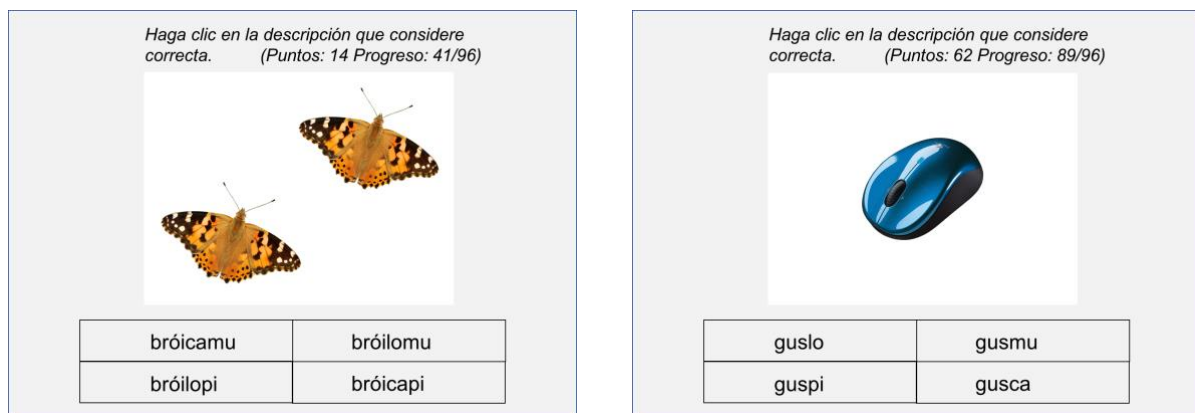
Again, four short multiple-choice tests were used to detect and filter out any non-native speakers. These were adapted to Spanish, but still tested advanced knowledge of one idiom and spotting of small grammatical errors as with the Turkish version (the tests can be found in the Appendix to this thesis). The adaptation to the Spanish language again happened with the help of native speakers and we ensured that googling would not provide easy cues to the correct answer. One of the Spanish informants was also contacted after completion of the experiment in order to verify my Spanish-English translations of the questionnaire responses. This was necessary to ensure correct interpretation of participants' strategies and to verify the exclusion of those participants who did not meet the participation criteria.

### 5.2.3 Procedure

The procedure was the same as in the previous experiments presented in **Chapters 3 and 4, Figure 1** showing a trial in the agglutinating and a trial in the fusional condition. Participants again received 1 point for each correct answer and their score was tracked throughout the experiment, consisting of 96 trials which each showed a unique picture to avoid any image or stem repetition. After the experiment, participants

completed the final questionnaire which was equivalent to those used with the L1 English and the L1 Turkish participants.

Prior to the actual experiment, participants completed two forced-choice tests with four options each which were used to filter out any non-native speakers. Responses in the final questionnaire of the L1 English experiments suggested that some submissions might have been carried out by bots. Since the present experiment was again conducted on Amazon Mechanical Turk, the native speaker tests should help eliminate these types of submissions as well. After completion of the 96 trials and before the final questionnaire, participants needed to complete two further tests. If a participant failed to provide the correct answer in one of the first two tests, a message appeared on the screen stating that the selected answer was incorrect and that the participant was not able to continue the experiment. If an incorrect answer was selected in the final two tests, no message appeared and the participant was able to complete the experiment, but their data was excluded. The experiment lasted for approximately 25 minutes.



**Figure 1:** Example trials in the agglutinating condition (left) and in the fusional condition (right) in Experiment 5, showing an animate plural trial and an inanimate singular trial, respectively.

## 5.3 Results

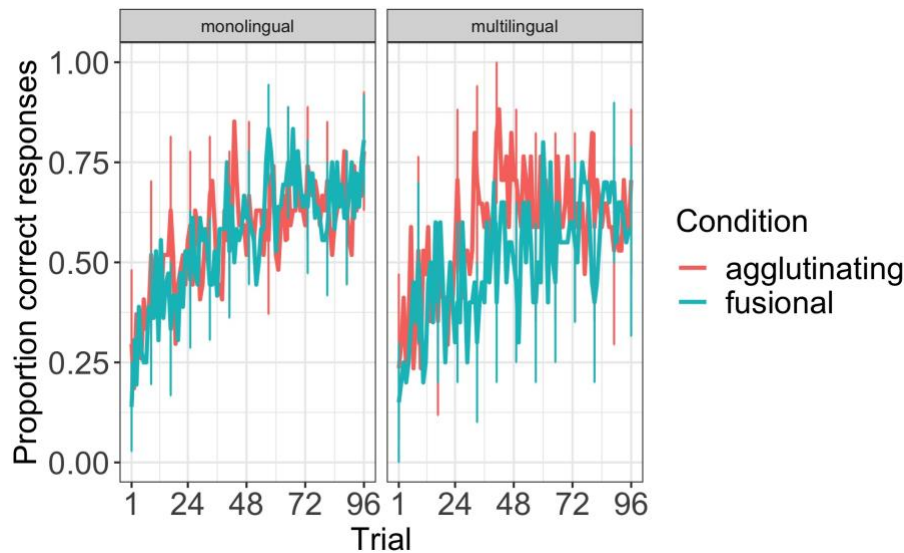
### 5.3.1 Overall accuracy

The experiment was again used to compare performance across conditions over time. As in the previous experiments, participants necessarily needed to take a guess during the initial trials due to the adaptive tracking setup. Therefore, performance was

expected to be similar across conditions early on. This time, there are two main outcomes with respect to a divergence of performance between conditions over time which can be motivated by previous research. First, Spanish-speaking participants might have a bias for transparent morphemes, given the strong claims in the literature that transparent forms are easier to learn. Note that if morphological technique can be transferred similar to abstract categories, then Spanish-speaking participants might have a bias for morphemes which are non-transparent in their form-meaning mapping due to Spanish being a largely fusional language. However, this option was considered less plausible. Studies have also suggested that transparent structures are learned better than non-transparent ones despite the occurrence of non-transparent forms in the learners' L1 (Igartua, 2015). Second, Spanish-speaking participants might overall perform better in the experiment than English-speaking participants due to their more profound experience with inflectional morphology from their native language(s). Thus, the learning curve is expected to be steeper and to increase earlier in the agglutinating compared to the fusional condition, and potentially, learning rates overall may be higher in this experiment than in Experiment 1 from **Chapter 3**. Note that this direct comparison is possible since both L1 groups were tested via Amazon Mechanical Turk.

**Figure 2** shows mean accuracy by trial across conditions for the monolingual and the multilingual group. The distribution of participants across conditions and groups was as follows. 27 monolingual participants were assigned to the agglutinating condition, 36 to the fusional condition. Of the multilingual group, 17 participants were assigned to the agglutinating condition and 20 to the fusional condition. As expected and observed in the previous experiments, participants started at around chance level (25% correct due to random selection among 4 options) and generally improved over time. There is a tendency for participants in the multilingual group to differ more between conditions than in the monolingual group. Participants again reached a relatively high level of accuracy overall by the end of the experiment. Mean accuracy across all trials for the monolingual group was 0.57 (SD=0.28) in the agglutinating condition and 0.56 (SD=0.21) in the fusional condition. For the multilingual group, mean accuracy across all trials was 0.58 (SD=0.28) in the agglutinating condition and 0.50 (SD=0.21) in the fusional condition. Across the final 10 trials, mean accuracy for the monolingual group was 0.66 (SD=0.33) in the agglutinating condition and 0.69 (SD=0.29) in the fusional condition. For the multilingual group, mean accuracy across

the final 10 trials was 0.59 (SD=0.39) in the agglutinating condition and 0.62 (SD=0.29) in the fusional condition. There is a very slight tendency for participants in both groups to perform better in the fusional compared to the agglutinating condition in the final 10 trials. Overall, the mean learning success of the L1 Spanish group appears to be slightly lower than that of the English and Turkish native speaker groups. This impression will be tested in a final model at the end of this chapter.



**Figure 2:** Mean accuracy by trial number, condition and number of L1s (monolingual vs. non-monolingual) in Experiment 5 with Spanish native speakers. Error bars indicate 95% bootstrapped confidence intervals.

To test whether the rate of improvement differed between conditions I fitted a logistic polynomial mixed-effects regression model as with the experiments reported in **Chapters 3** and **4**. The learning curves were modelled using second-order orthogonal polynomials, predicting correct answer by condition, trial, number of native languages and their respective interactions. Since the polynomials were orthogonal, the model intercept refers to the overall average performance.<sup>18</sup> Trial was replaced by a linear

<sup>18</sup> All models reported in this chapter were run using the package lme4 (Bates, 2015) in R (R Core Team, 2020). The specification of the polynomial models generally follows the procedure outlined in Mirman (2014, pp.44-60). I begin with a base model including all experimental predictors as well as a linear time term ot1. A subsequent model includes the quadratic time term ot2:

Model 1: `answer_correct ~ condition * number_of_L1s * ot1 + (1 + ot1 | participant) + (1 | item)`  
 Model 2: `answer_correct ~ condition * number_of_L1s * (ot1 + ot2) + (1 + (ot1 + ot2) | participant) + (1 | item)`

Models are compared using a Likelihood Ratio Test to evaluate whether the more complex model is justified by the data. Unless otherwise noted, models included random by-participant and by-

and a quadratic time term; the time terms were added sequentially and the fit of the resulting two models (linear and polynomial) was compared using a Likelihood Ratio Test. Random by-item intercepts were removed from both models due to convergence failure.

The inclusion of the quadratic time term significantly improved model fit ( $X^2=47.05$ ,  $p<0.001$ ). The polynomial model revealed a significant effect of the first time term ( $b=7.96$ ,  $SE=1.04$ ,  $p<0.001$ ), and of the second time term ( $b=-1.54$ ,  $SE=0.43$ ,  $p<0.001$ ). These results indicate that participants improved their accuracy over the course of the experiment and that their performance flattened out towards the end. This is the same pattern that was observed in all of the previous experiments as well.

There was no significant effect of condition ( $b=0.23$ ,  $SE=0.19$ ,  $p=0.216$ ; note that this indicates the effect of condition when the two time terms are 0, which was at the midpoint of the experiment). This result means that participants did not perform better in one condition over the other. Again, this is the same tendency as observed in the previous experiments with different native speaker groups. Further, there was no significant effect of the number of L1s ( $b=0.12$ ,  $SE=0.19$ ,  $p=0.521$ ), which suggests that monolingual and non-monolingual participants performed similarly in the experiment.

There was no significant interaction effect between condition and the first time term ( $b=0.14$ ,  $SE=1.00$ ,  $p=0.889$ ), but there was a significant interaction effect between condition and the second time term ( $b=-0.79$ ,  $SE=0.38$ ,  $p=0.038$ ). These results indicate that there was no difference in acquisition rates across conditions over time, but that the learning curve flattened out more in the agglutinating condition. The model further indicates that this effect is driven by participants in the multilingual group, since the effect of the interaction between condition, number of L1s and the second time term was significant ( $b=0.81$ ,  $SE=0.38$ ,  $p=0.032$ ), also shown in **Figure 2**. The learning curve of the group of multilingual participants in the agglutinating condition flattened out more over time compared to the monolingual group. The effect of the interaction between condition, number of L1s and the first time term was not significant

---

item (picture) intercepts, and by-participant slopes for the time terms. The predictors of condition (agglutinating vs. fusional) and number of L1s (monolingual vs. multilingual) were effects-coded with agglutinating and monolingual compared to the grand average. All models used the optimizer 'bobyqa'.

( $b=-0.02$ ,  $SE=0.99$ ,  $p=0.98$ ), indicating that the learning curves did not differ significantly between monolingual and multilingual participants in the agglutinating condition. None of the other interaction effects involving number of L1s in the model were significant. Full model results are shown in **Table 2** in the Appendix to this chapter. Overall, there is yet again no indication of a learnability advantage of agglutinating over fusional morphology.

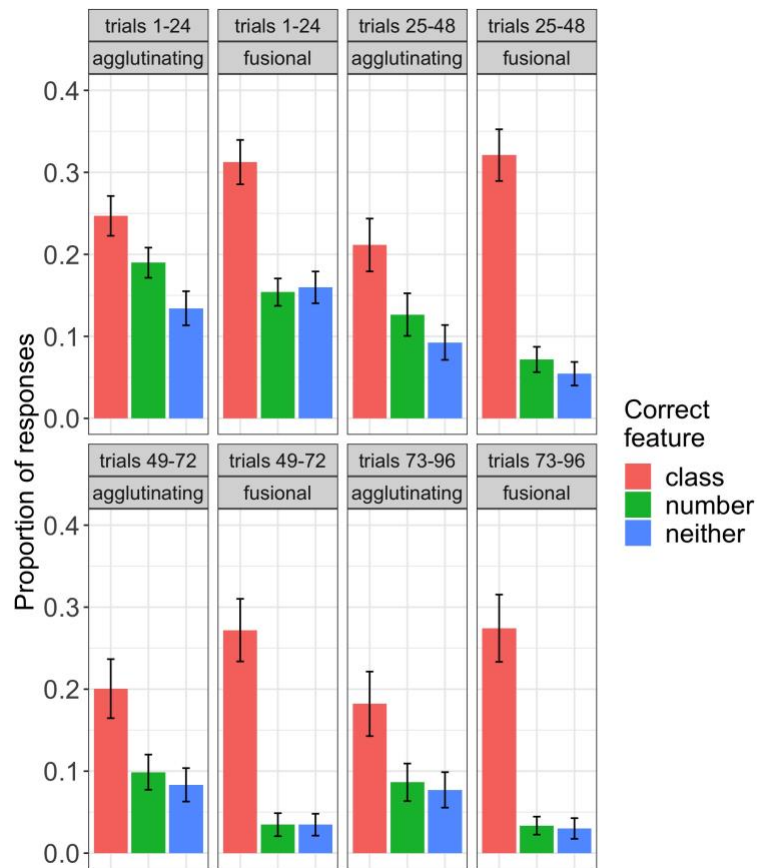
### 5.3.2 Discussion

The results for overall accuracy in the present experiment with L1 Spanish speakers are strikingly similar to those of all previous experiments with L1 English- and L1 Turkish-speaking participants. Learning improves over time, but there is no significant effect of condition on learnability, and therefore no general advantage of the agglutinating over the fusional condition. This result contradicts the third prediction, which finds strong support in previous research (Aksu-Koç & Slobin, 1985; Bittner et al., 2003; Brown, 1976; Don, 2017; Hengeveld & Leufkens, 2018; Igartua, 2015; Penke, 2012). Reversely, L1 Spanish-speaking participants do not learn better in the fusional than in the agglutinating condition either, but this second prediction appeared less supported by existing research.

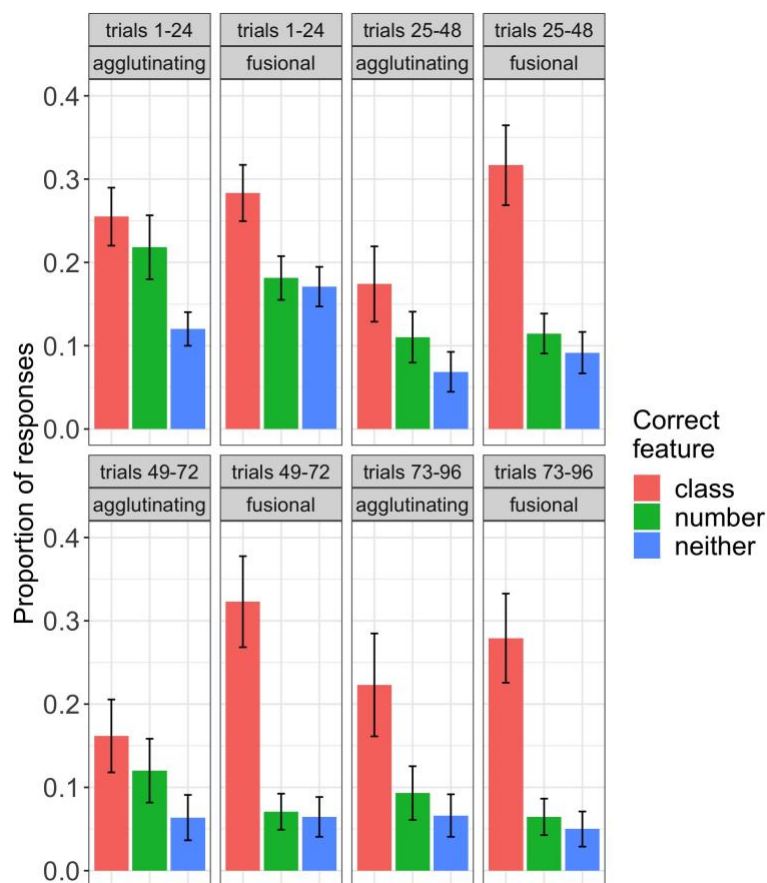
The first, and more general prediction, was that Spanish-speaking participants might learn more successfully than English-speaking participants in both experimental conditions due to Spanish being inflectionally richer than English. Since the L1 Spanish group was tested on Amazon Mechanical Turk like the L1 English group, any potential differences between them cannot be due to different testing platforms. However, it appears as if the Spanish participant group performed slightly worse than both other L1 groups (Experiment 1 with the L1 English group and Experiment 4 with the L1 Turkish group) with regards to mean accuracy per condition and mean accuracy in the final 10 trials. In comparison, the L1 Turkish group showed more similar mean accuracy values to the L1 English group in Experiment 1 despite having been tested on a different platform. This impression of a general learning difference between L1 groups will be tested in a final model in which all three groups of native speakers will be compared. Before this, the pattern of errors in Experiment 5 with L1 Spanish-speakers will be analysed.

### 5.3.3 Analysis of error patterns

In this experiment, participants' native language was overall fusional, which could theoretically have led to a reduced bias for transparent structures, and therefore to a lower amount of CLASS correct only responses compared to the previous L1 participant groups. However, this was considered less plausible given common claims about learning preferences for transparent structures. Additionally, Spanish marks an ANIMACY distinction to some extent. Moreover, some multilingual participants were native speakers of Basque, which shows an ANIMACY-based distinction in the marking of local cases of noun phrases (Hualde & Ortiz de Urbina, 2003). The opposition between the fusional L1 and the presence of ANIMACY as a grammatical feature in the L1 would lead to opposite predictions with regards to the amount of CLASS correct only responses if one assumes transfer effects – if the fusional nature of the L1 has a stronger impact, then this type of error should be less prominent since participants would not try to enforce transparent morphemes; if the feature of ANIMACY has a stronger effect, then we might see CLASS correct only responses more frequently since participants might be more attentive to CLASS. If the distribution of error types is in the range of the L1 English and L1 Turkish participant groups, then neither characteristic of the L1 appears to have a strong impact. This would suggest that any biases observed are more general and less dependent on any specific L1.



**Figure 3:** Proportion of the three possible types of error (CLASS correct only, NUMBER correct only, none correct) in the monolingual group, split by condition and across four sections of Experiment 5. Note that only the subset of incorrect answers is shown.

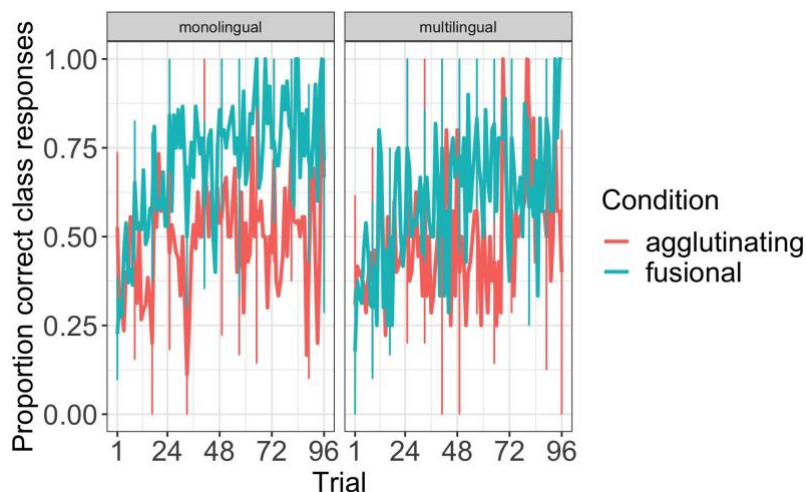


**Figure 4:** Proportion of the three possible types of error (CLASS correct only, NUMBER correct only, none correct) in the multilingual group, split by condition and across four sections of Experiment 5. Note that only the subset of incorrect answers is shown.

The distribution of error types is shown in **Figures 3** and **4**. It resembles that of previous experiments in the sense that CLASS correct only responses were the most frequent type of error in each of the four sections of the experiment, independent of condition and whether participants were monolingual or not. This means that participants were dominantly selecting answers that indicated the correct value for CLASS, but the incorrect value for NUMBER (for instance choosing animate singular in an animate plural trial). **Figures 3** and **4** also reveal a tendency of CLASS correct only responses to be more frequent in the fusional than in the agglutinating condition. In the latter, CLASS correct only responses still appear to dominate, but there seem to be more NUMBER correct only responses in comparison to the fusional condition. Nevertheless, CLASS correct only responses remain dominant throughout the experiment. **Figure 5** singles out CLASS correct only responses and displays their proportion among incorrect answers across trials in both conditions for monolingual

and multilingual participants. I again assessed the difference between conditions and monolingual/multilingual groups in choosing CLASS correct only responses using a logistic polynomial mixed-effects regression model fit to the subset of the data with incorrect answers. The dependent variable was CLASS correct. I again used second-order orthogonal polynomials with fixed effects of condition, number of L1s, trial number (replaced by a linear and a quadratic time term) and their interactions. Condition and number of L1s were effects-coded whereby agglutinating and monolingual were compared to the grand average. Random by-item intercepts were removed from the polynomial model due to convergence failure.

The inclusion of the quadratic time term significantly improved model fit ( $X^2=12.98$ ,  $p=0.043$ ). The polynomial model revealed a significant effect of the first time term ( $b=5.07$ ,  $SE=0.84$ ,  $p<0.001$ ) and of the second time term ( $b=-1.05$ ,  $SE=0.46$ ,  $p=0.022$ ). This indicates that among incorrect responses, CLASS correct only responses increased over time and that the curve of this type of response flattened out towards the end of the experiment. The effect of condition was also significant ( $b=-0.57$ ,  $SE=0.17$ ,  $p<0.001$ ), indicating that participants in the agglutinating condition produced fewer CLASS correct only responses than those in the fusional condition at the midpoint of the experiment. The interaction between condition and the first time term was also significant ( $b=-2.77$ ,  $SE=0.83$ ,  $p<0.001$ ), and so was the interaction between condition and the second time term ( $b=0.90$ ,  $SE=0.44$ ,  $p=0.043$ ). These results suggest that CLASS correct only responses decreased more rapidly over time in the agglutinating than in the fusional condition, but the curve flattened out less in the agglutinating condition compared to the fusional condition. The effect of the number of L1s was not significant ( $b=0.22$ ,  $SE=0.17$ ,  $p=0.206$ ), indicating that monolingual and multilingual participants selected CLASS correct only answers to a similar extent at the midpoint of the experiment. None of the other interaction effects involving number of L1s were significant either, suggesting that the choice of CLASS correct only answers is independent of how many native languages participants had. Full model results are shown in **Table 3** in the Appendix to this chapter.



**Figure 5:** Mean proportion of incorrect responses in which only CLASS was correct, by trial number, number of L1s and condition. Error bars indicate 95% bootstrapped confidence intervals. Only the proportion among incorrect answers is shown.

### 5.3.4 Discussion

In this experiment, different outcomes with regards to a focus on CLASS were plausible. On the one hand, the fact that Spanish displays many fusional characteristics could have meant that this reduces the slight transparency bias which was observed for the L1 English and L1 Turkish participants in the analysis of error types. If fusional morphology can exercise transfer effects from the L1 onto the L2, then we would expect Spanish-speaking participants to aim less at making fusional morphemes transparent. Remember that a stronger focus on CLASS, in other words a larger proportion of CLASS correct only responses, in the fusional compared to the agglutinating condition reflects participants' strategy of making the fusional morphemes transparent – as was observed earlier, participants were trying to enforce a four-way distinction of CLASS onto the morphemes, ignoring NUMBER altogether. This means that the fusional morphemes were understood as expressing only one feature, namely CLASS. On the other hand, Spanish, as well as Basque, marks an ANIMACY distinction grammatically. If knowledge of ANIMACY as a feature was exercising transfer effects onto the L2, then we would expect Spanish-speaking participants to explore ANIMACY more and to produce more CLASS correct only responses. This was considered the more plausible prediction. As shown in the previous section, Spanish-speaking participants display a focus on CLASS very much like the English- and

Turkish-speaking participants, namely more in the fusional than in the agglutinating condition.

These results have several implications. First, they suggest that native speakers of Spanish do not have a noticeable preference for fusional structures; otherwise they should have been better at learning both CLASS and NUMBER, especially in the fusional condition, than the other participant groups. Instead, it appears as if error patterns are similar across all experiments. Second, based on comparing figures across experiments, L1 Spanish-speaking participants do not show a considerably stronger focus on CLASS than participants with other L1s. This impression will be tested in a comprehensive model further below, but it suggests that ANIMACY as a feature may not exert strong transfer effects in this experiment either. Instead, the error patterns of the current experiment are not much different to any previously observed ones, and in this way, they support the interpretation that participants generally have a bias – albeit arguably weak – for morphemes to be transparent, leading them to explore possible categorisations within the CLASS feature, and ignoring the NUMBER feature. It needs to be emphasised, however, that despite the apparent transparency bias in selecting answers, it does not translate into better learning in the agglutinating over the fusional condition, and therefore appears to be weak at best. What can be noted, however, is that participants seem to be learning very similarly despite having different native languages. Since Spanish-speaking participants reported similar learning strategies to the Turkish-speaking participants in the final questionnaire, their answers will not be discussed in detail. The only point to note is that some participants explicitly described considering ANIMACY as a feature expressed in the stimuli.

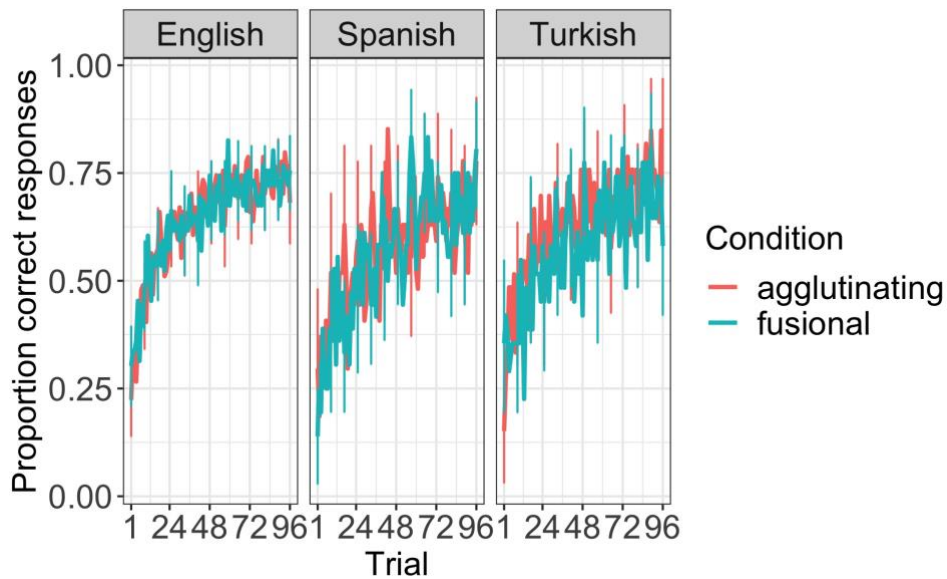
#### **5.4 Comparison of all L1 groups: English, Turkish and Spanish native speakers**

As a final step of investigating the role of the native language, more precisely of the morphological structure of the native language, in the acquisition of L2 morphology, the results from all three groups of participants will be compared. Performance appeared relatively similar across all experiments so far, in that there was no effect of transparency on overall learning, but that participants exhibited a bias for transparent structures when comparing the types of errors made. Nevertheless, it is possible that depending on their L1, participants learned better overall or better in one condition

over the other. For instance, it appeared as if L1 Spanish-speaking participants were learning slightly worse on average than the other two groups. It is also possible that speaking a particular L1 led some participants to make different types of errors than other participants in an overall comparison. In order to investigate this, the data from Experiment 1 in **Chapter 3**, from Experiment 4 in **Chapter 4** and from Experiment 5 in the present chapter were pooled together. In all three data sets, only the monolingual participants were included to ensure comparability of the data. The resulting dataset consisted of 180 native speakers of English (94 in the agglutinating and 86 in the fusional condition), 64 native speakers of Turkish (33 in the agglutinating and 31 in the fusional condition) and 63 native speakers of Spanish (27 in the agglutinating and 36 in the fusional condition). The analysis followed the same structure as the previous analyses.

#### **5.4.1 Overall accuracy**

At first, overall performance across conditions over time was compared between native speakers of English, Turkish and Spanish. **Figure 6** shows mean accuracy by trial across conditions for all three L1 groups. The learning curves between the three groups appear similar with regards to an increase of accuracy rates over time and a similarity of performance across both conditions. There appear to be slight differences in the rate of accuracy between L1 groups, with the L1 English group showing a slightly higher rate and more consistency. However, this group contains more data points than the other groups.



**Figure 6:** Mean accuracy by trial number, condition and L1 in the combined data from Experiments 1, 4 and 5. Only data from monolingual participants is shown. Error bars indicate 95% bootstrapped confidence intervals.

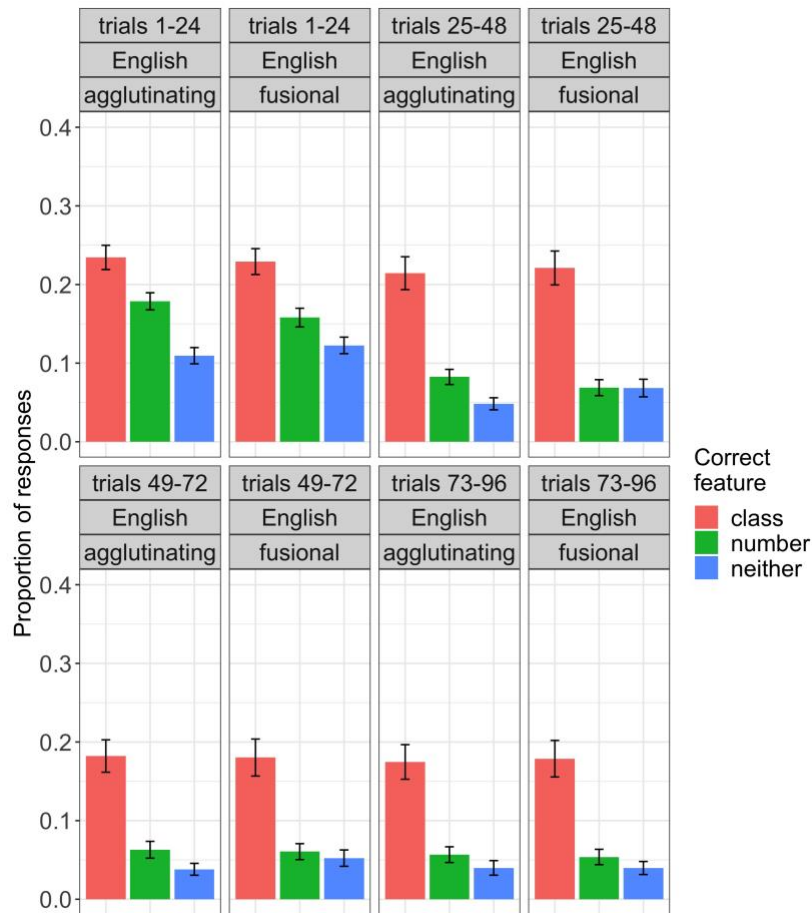
To test whether there were any differences in learning rates across groups, I again fitted a logistic polynomial mixed-effects regression model. The learning curves were modelled using second-order orthogonal polynomials, predicting correct answer by condition, trial, L1 and their respective interactions. The predictors of condition and L1 were effects-coded with agglutinating, Turkish and Spanish being compared to the grand average.

The inclusion of the quadratic time term significantly improved model fit ( $X^2=160.69$ ,  $p<0.001$ ). As with the previous analyses, the polynomial model revealed a significant effect of the first time term ( $b=8.65$ ,  $SE=0.62$ ,  $p<0.001$ ), and of the second time term ( $b=-1.60$ ,  $SE=0.30$ ,  $p<0.001$ ). The effect of condition was not significant ( $b=0.11$ ,  $SE=0.13$ ,  $p=0.394$ ), indicating that learning was similar in both conditions at the midpoint of the experiment. The effects of L1 were not significant either ( $b=-0.11$ ,  $SE=0.19$ ,  $p=0.546$  for Turkish and  $b=-0.24$ ,  $SE=0.19$ ,  $p=0.218$  for Spanish, respectively). This suggests that the performance of native speakers of Turkish and Spanish was not different from the grand average in the middle of the experiment. No other effects in the model were significant. Importantly, this means that none of the interaction effects involving condition, trial, L1 or all of these together, led to a learning performance that was different from the grand average. In other words, the increase and flattening out of the learning curve over time and the lack of an effect of transparency on learning is independent of the native language of the participants. Full

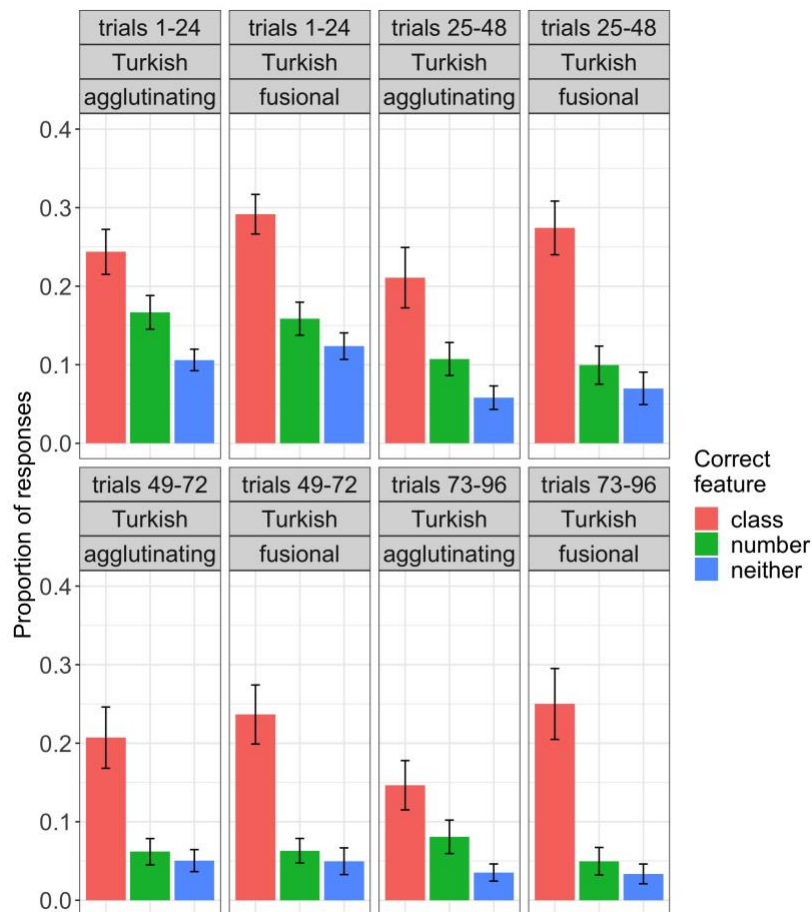
model results are shown in **Table 4** in the Appendix to this chapter. Overall, the results from the comprehensive model suggest that learning the morphology of the artificial language follows a similar path across groups of different L1s.

#### **5.4.2 Analysis of error patterns**

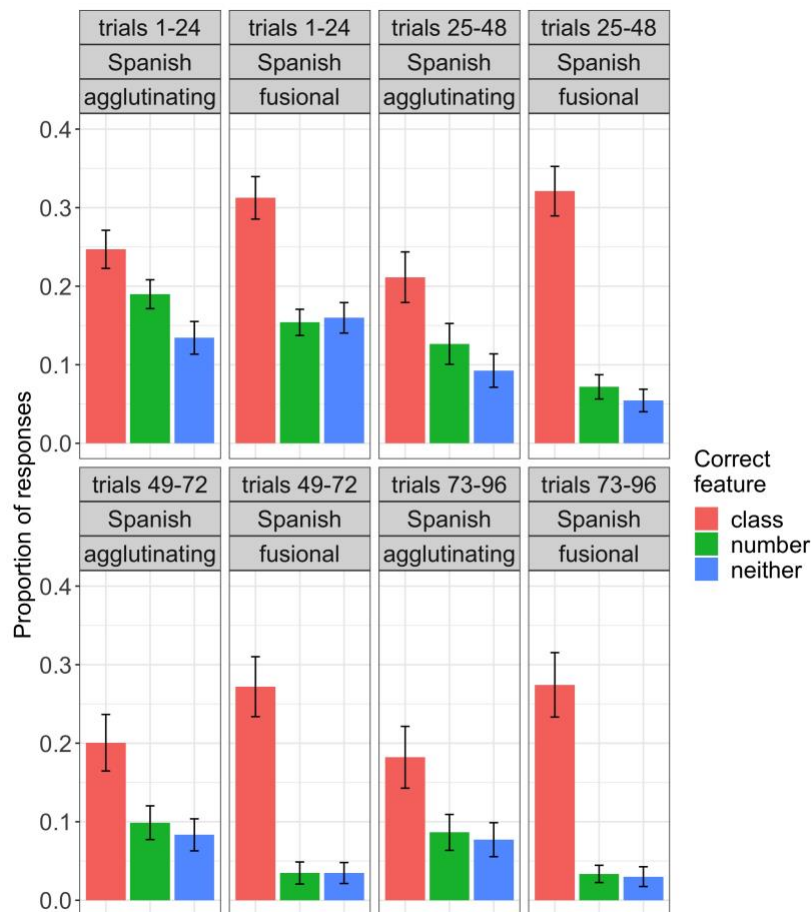
In addition to overall accuracy, the proportion of each type of errors may also have been different depending on the L1 of the participants, although all three groups had a tendency to produce more CLASS correct only responses than other error types. Specifically, I will test whether the slight bias for transparency that was observed in the analyses of error patterns so far is of a different magnitude depending on L1. Given previous research, this bias should be stronger in the L1 Turkish group and weaker in the L1 Spanish group compared to the average performance if we assume that the morphology of the L1 plays a role. When classifying incorrect answers as CLASS correct only, NUMBER correct only, or none correct, a few patterns are noteworthy. **Figures 7-9** show their distribution across four sections in the experiment per condition, separated by L1 group. Overall, CLASS correct only responses are the most dominant type of error in each panel, that is for every L1 group and for both conditions across the four sections of the experiment. It can also be noted that this error type appears to be more frequent in the fusional than in the agglutinating condition within each L1 group and each experimental stage. These distributions were essentially found in each of the separate experiments. In addition, there also appears to be a slight tendency for the difference in CLASS correct only responses between both conditions to be larger for the Spanish than the other L1 groups. These impressions were tested in another comprehensive model of the combined data set, following the earlier modelling approach.



**Figure 7:** Proportion of the three possible types of error (CLASS correct only, NUMBER correct only, none correct) per condition in the L1 English group, split up across four sections of the experiment. Note that only the subset of incorrect answers is shown.

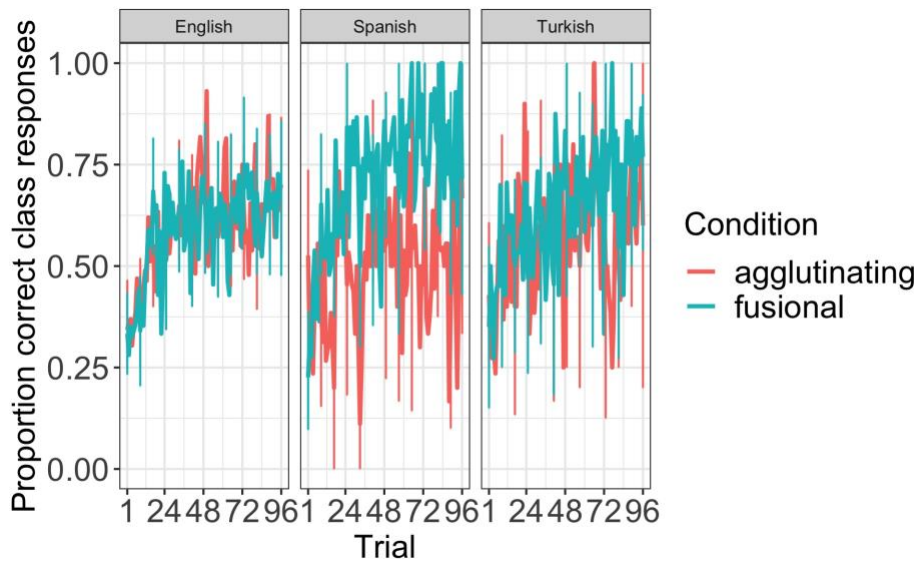


**Figure 8:** Proportion of the three possible types of error (CLASS correct only, NUMBER correct only, none correct) per condition in the L1 Turkish group, split up across four sections of the experiment. Note that only the subset of incorrect answers is shown.



**Figure 9:** Proportion of the three possible types of error (CLASS correct only, NUMBER correct only, none correct) per condition in the L1 Spanish group, split up across four sections of the experiment. Note that only the subset of incorrect answers is shown.

**Figure 10** singles out CLASS correct only responses and displays their proportion among incorrect answers across trials in both conditions for each L1 group. The difference in CLASS correct only responses between conditions and L1 groups was again assessed using a logistic polynomial mixed-effects regression model fit to the subset of the data with incorrect answers. The dependent variable was CLASS correct. I again used second-order orthogonal polynomials with fixed effects of condition, L1, trial number (replaced by a linear and a quadratic time term) and their interactions. Condition and L1 were effects-coded whereby agglutinating, Turkish and Spanish were compared to the grand average.



**Figure 10:** Mean proportion of incorrect responses in which only CLASS was correct, by trial number, L1 and condition in the combined data from Experiments 1, 4 and 5. Error bars indicate 95% bootstrapped confidence intervals. Only the proportion among incorrect answers is shown.

The inclusion of the quadratic time term significantly improved model fit ( $X^2= 50.96$ ,  $p<0.001$ ). The polynomial model revealed a significant effect of the first time term ( $b=4.55$ ,  $SE=0.51$ ,  $p<0.001$ ) and of the second time term ( $b=-1.21$ ,  $SE=0.32$ ,  $p<0.001$ ). Also the effect of condition was significant ( $b=-0.378$ ,  $SE=0.10$ ,  $p<0.001$ ), indicating that participants in the agglutinating condition produced fewer CLASS correct only responses than those in the fusional condition at the midpoint of the experiment. The interaction between condition and the first time term was also significant ( $b=-1.74$ ,  $SE=0.51$ ,  $p<0.001$ ), but not the interaction between condition and the second time term ( $b=0.15$ ,  $SE=0.32$ ,  $p=0.626$ ). This indicates that CLASS correct only responses decreased more rapidly over time in the agglutinating compared to the fusional condition, but that the flattening of the curve of this type of error was similar between conditions towards the end of the experiment. This pattern of results largely reflects what has been found in the analyses of the individual experiments earlier.

There was no significant effect of L1 ( $b=-0.07$ ,  $SE=0.15$ ,  $p=0.632$  for Turkish and  $b=0.23$ ,  $SE=0.15$ ,  $p=0.141$  for Spanish, respectively), which suggests that neither of the L1s had a rate of CLASS correct only responses that was markedly different from the grand average at the midpoint of the experiment. However, the interaction between condition and the Spanish L1 was significant ( $b=-0.31$ ,  $SE=0.15$ ,  $p=0.043$ ). This indicates that the difference between conditions in CLASS correct only responses is

largest for the group of Spanish native speakers. Similarly, also the interaction between condition, the Spanish L1 group and the first time term was significant ( $b=-1.52$ ,  $SE=0.76$ ,  $p=0.045$ ), indicating that across trials, CLASS correct only responses decreased more rapidly in the agglutinating compared to the fusional condition for the group of Spanish native speakers. No other effects were significant and full model results can be found in **Table 5** in the Appendix to this chapter.

### 5.4.3 Discussion

The analysis of the combined data sets from the experiments with three different native languages, representing the isolating, agglutinating and fusional types, has shown that learning rates increase over time, but essentially develop very similarly, independent of participants' mother tongue. Three main conclusions can be drawn from this. First, the model again suggests that there is no benefit of transparent over non-transparent morphemes across the experiments presented in this thesis. Second, being a native speaker of an inflectional language (Turkish and Spanish) does not result in different learning patterns in one condition over the other. Similarly, being a native speaker of a (non-)transparent language does not lead to better learning of (non-)transparent morphology in the experiments presented here. This indicates that experience with a particular morphological technique in the mother tongue does not result in a predisposition to learn this type of morphology more easily in an L2. Put slightly differently, the effect of transparency on learnability that is suggested in the literature is not confirmed even if participants speak a highly transparent L1.

Concerning the distribution of CLASS correct only responses across conditions and L1 groups, the analysis of the combined data largely confirmed the previously identified pattern that this type of error is more common in the fusional than in the agglutinating condition. This was interpreted as a slight bias for transparent structures, apparent through participants' exploration of a 4-way distinction based on CLASS, which allows them to ignore the feature of NUMBER and renders the fusional morphemes transparent.

However, the model also revealed that the difference in the occurrence of this error type between the agglutinating and the fusional condition was largest for the L1 Spanish group, a tendency which is confirmed in **Figure 6**. This illustration of error

types also suggests that NUMBER correct only responses occurred more often in the agglutinating condition in the Spanish L1 group than in the agglutinating condition in other L1 groups. It might be the case that Spanish native speakers were better able to learn both the CLASS and the NUMBER feature in the agglutinating condition compared to native speakers of English and Turkish. This may mean that L1 speakers of Spanish also have a bias for transparent structures, and that at the same time, they were more successful in considering a second feature in the agglutinating condition than the other participant groups. Despite this, CLASS correct only responses still dominate even in the L1 Spanish group, again revealing a more pronounced focus on CLASS in general. This focus may partly be driven by the fact that Spanish-speaking participants were familiar with ANIMACY distinctions through their L1, leading them to explore CLASS more, especially in the fusional condition. Concerning this group of participants, it is difficult to tell whether their erroneous responses are driven more by a bias for transparency or by transfer of ANIMACY distinctions from the L1 based on the data at hand.

However, the model on the combined data revealed that overall accuracy rates are similar across L1s and the analyses presented in **Chapter 3** have demonstrated that segmentation of morphemes happened successfully in Experiment 1 with L1 English speakers. The fact that Turkish- and Spanish-speaking participants have a similar overall accuracy rate to that of the English-speaking participants suggests that they may have segmented the grammatical morphemes to a similar extent. In light of this, it appears more plausible that the performance of Spanish-speaking participants in the agglutinating condition is driven by a bias for transparent structures: assuming that each morpheme only expresses one meaning and having identified two grammatical morphemes per word will lead participants to search for a second feature that is expressed by the second morpheme. Knowledge of ANIMACY alone cannot explain well why this group of participants displays a more similar learning of both features in the agglutinating condition. This lends further support to the interpretation that the level of transparency of the L1 is not a predictor of learning ease of transparent morphology in an L2, since Spanish is largely fusional. Finally, the group of Spanish-speaking participants again shows that if a transparency bias exists, it is still similarly weak as that of the other L1 participant groups, since L1 Spanish-speaking participants did not show any difference in overall accuracy rates across conditions.

## 5.5 General Discussion

Transparency has been claimed to be an important predictor of acquisition ease during both L1 and L2 learning for several decades (Brown, 1973; Goldschneider & DeKeyser, 2001; Igartua, 2015; Narasimhan & Gullberg, 2011; Penke, 2012). However, in many studies of natural languages, transparency is correlated with further characteristics of morphemes in paradigms. In the series of experiments presented in this thesis, transparency was singled out and tested largely in isolation by means of an artificial language learning paradigm. While native speakers of English and Turkish did not display the suggested effect of a transparency benefit, the group of Spanish native speakers was used to complement the test of the interplay of inflectional richness, agglutinating and fusional morphology in the L1 on the one hand and transparent and non-transparent morphology in the L2 on the other hand. Several predictions were made in light of previous research.

On the whole, each of the predictions formulated in the beginning is either not or only weakly confirmed. The first prediction was that native speakers of Spanish will be better at learning inflectional morphology in general, which should result in higher overall accuracy rates compared to the results from L1 English participants. However, the final model on the combined data reveals no difference across groups. In a similar vein, the fact that all three L1 groups show a similar overall accuracy rate suggests that participants segment the grammatical morphemes similarly well, along the lines of what has been established through a comparison of Experiments 1 and 2 in **Chapter 3**. There, it was shown that an additional colour aid to highlight morpheme boundaries did not yield higher scores, suggesting that participants segment the morphemes successfully in the absence of colour. This finding contradicts the assumption that an inflectionally rich L1 helps with the segmentation of morphemes in the input and results in better learning. While it is certainly possible that having an inflectionally rich L1 helps with the online processing and detection of L2 morphemes, the study is neither in a position to comment on online processing during the task, nor do the offline results for overall accuracy appear in any way different between the three L1 groups in either condition. This suggests that offline acquisition of morphology is not significantly impacted by the inflectional richness of the participants' mother tongue in this task.

Another possibility of an impact from the L1 is that experience with a particular morphological technique (agglutination, fusion) will make it easier to acquire this type of morphology in an L2, although this was considered a less likely scenario. Such an effect should have again manifested itself in differences in overall accuracy in one of the two conditions depending on the L1. For the present experiment, the corresponding situation would have been that Spanish L1 speakers should find it easier to acquire the artificial language in the fusional condition if L1 experience overrides any general biases for transparent structures. However, the analysis shows that native speakers of Spanish learn similarly well in both conditions, which is exactly what has been found for native speakers of English and Turkish. On the other hand, if learning is determined less by features of the L1 and more by cognitive preferences for transparent structures (second prediction), then Spanish speakers should have been more successful in acquiring the language in the agglutinating condition. However, as with the earlier experiments, this is not the case either.

The only area where native speakers of Spanish differ from the grand average performance is that their focus on CLASS is more markedly different between the agglutinating and the fusional condition, a pattern that is revealed when investigating the types of errors made (third prediction). It appears that this group has learned NUMBER and CLASS in more similar ways in the agglutinating condition, but shows a stronger focus on CLASS in the fusional condition than the previous participant groups. It may be that this is linked to the fact that Spanish nouns mark NUMBER (only the plural is marked overtly), but also contain indicators for GENDER. Both are often separate elements, although GENDER markers are not necessarily morphemes, but can be regarded as part of the stem (Fábregas, 2018). Examples of nouns expressing GENDER and NUMBER are *casa-casas* (house-houses) and *perro-perros* (dog-dogs). This may have helped participants to look for similar indicators in the suffixes of the stimuli words, resulting in better acquisition of the NUMBER feature. For the main research question, this is only marginally relevant, however. L1 knowledge may have helped in noticing the grammatically relevant categories, but this has not resulted in better learning in one condition over the other.

Relatedly, Spanish also uses DOM, which is considerably dependent on ANIMACY distinctions (human animate, non-human animate, inanimate). Basque, which some of the participants spoke as another L1, also shows a restricted grammatical distinction based on ANIMACY. Spanish speakers' larger focus on CLASS in the fusional

condition compared to the agglutinating condition may have partly resulted from the fact that they are aware of ANIMACY due to their L1, and therefore explore possible categorisations of the images along this dimension in more depth. However, it appears that at the same time, participants were exploring CLASS and thereby ignoring NUMBER similarly to the other L1 groups, since CLASS correct only responses were the most dominant error type. This suggests that knowledge of fusional morphology from the L1 did not help them learn better in the fusional condition than other participants, since one feature was learned better than the other. Thus, transfer of morphological structures from the L1 to the L2 seems to be weak at best in this participant group.

Returning to the suggested effect of morpheme transparency on learning (second prediction), the L1 Spanish participants are the third group tested in this thesis that does not display the alleged learning facilitation. Like Turkish, Spanish is rich in inflections, which was expected to help participants learn inflectional morphology better compared to the L1 English group (first prediction). In the case of L1 Turkish speakers, the prediction based on transparency and the prediction based on morphology dominant in the L1 lined up, which is why learning ease was expected to be larger in the agglutinating condition for these participants than for Spanish-speaking participants. The prediction was not confirmed, but the Turkish-speaking participants were tested on a different platform and might have displayed other chance idiosyncracies. Testing native speakers of Spanish therefore provided another opportunity to test the suggested effect. However, performance across groups is relatively stable. Second, it was assumed that the general lack of inflections in English may have obscured the transparency advantage. However, like the L1 Turkish group, the L1 Spanish group does not exhibit a preference for the agglutinating condition either. Despite the slight bias that is visible in the error patterns, whether morphemes are transparent or not does not appear to lead to any difference in learnability and thereby contradicts a common claim in the literature (Aksu-Koç & Slobin, 1985; Brown, 1976; Bittner et al., 2003; Hengeveld & Leufkens, 2018; Igartua, 2015). This is despite the fact that transparency was the main difference between the agglutinating and the fusional condition in the experiments conducted.

The amount of data gathered through the five individual experiments with three groups of L1 is substantial with a total of 644 participants. The artificial language used in the two conditions was also carefully controlled, using the same morphemes in each and drawing participants' attention to the grammatical morphemes rather than the

stems or to learning novel phonotactic patterns. In addition to this, two potential sources of learning difficulty in the agglutinating condition were tested, namely segmentation difficulty and word length. In two separate experiments (Experiments 2 and 3 in **Chapter 3**), it was found that neither of them impacts learning in the agglutinating condition negatively. In theory, this should have allowed the benefit of transparency to take hold, but it was not found in any of the experiments. Therefore, the likelihood of transparency facilitating morpheme acquisition is in even stronger doubt given the results from the L1 Spanish group. Instead, it appears much more likely that learnability differences between morphemes which are thought to stem from transparency are in fact driven by other factors. Among them are generalisability and frequency, both of which are a result of the compositional structure of agglutinating systems, and the lower number of individual morphemes required. How they may have been overlooked in favour of transparency in various studies will be discussed in **Chapter 6**.

## 5.6 Conclusion

The purpose of this chapter was to complement the research agenda of testing the role of transparency and inflectional richness in the L1 for the acquisition of morphology in an L2. It has frequently been claimed that transparent morphemes are easier to acquire than non-transparent ones. Since English is inflectionally poor, this was initially considered a potential reason why a learnability advantage of the agglutinating condition was not found. By testing native speakers of Turkish and Spanish, I was able to investigate whether the initial findings reported in **Chapter 3** were due to a lack of experience with inflections, and whether a particular type of morphology is needed in the L1 to lead to a substantial transparency bias. While it was found that this was not the case for Turkish native speakers, the reverse was not the case either for native speakers of Spanish, who did not learn better in the fusional condition, which matches the type of morphology dominant in their L1.

In the particular experiment conducted here, transfer of abstract categories like GENDER and ANIMACY also only happened to a small extent, visible only when looking at the error patterns. A final model on the combined data set of all comparable experiments with native speakers of English, Turkish and Spanish confirmed the earlier impression from the individual experiments that a learnability advantage of

transparent structures cannot be found. Indeed, this result proved to be independent of the participants' native language. These overall results are highly important, since the claim that transparent form-meaning mappings are either the default or easier to learn, or both, is very common in the literature. The next chapter will return to this claim and re-consider these earlier proposals in line with the present findings from artificial language learning experiments. Thereby, it will be illustrated how a more nuanced view may need to be adapted when trying to explain the benefits of transparent morphology.

## 5.7 Appendix

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	0.73	0.19	3.82	<0.001
condition1	0.23	0.19	1.24	0.216
ot1	7.96	1.04	7.64	<0.001
ot2	-1.54	0.43	-3.58	<0.001
lingual1	0.12	0.19	0.64	0.521
condition1:ot1	0.14	1.00	0.14	0.889
condition1:ot2	-0.79	0.38	-2.07	0.038
condition1:lingual1	-0.09	0.19	-0.46	0.647
ot1:lingual1	0.91	0.99	0.92	0.360
ot2:lingual1	0.37	0.38	0.98	0.328
condition1:ot1:lingual1	-0.02	0.99	-0.02	0.984
condition1:ot2:lingual1	0.81	0.38	2.14	0.032

**Table 2:** Results from Experiment 5 with Spanish native speakers, overall accuracy. The predictors of condition and number of L1s were effects-coded with agglutinating and monolingual being compared to the grand average.

	Estimate	Std. Error	z value	Pr(>  z )
(Intercept)	0.61	0.17	3.50	<0.001
condition1	-0.57	0.17	-3.33	<0.001
ot1	5.07	0.84	6.03	<0.001
ot2	-1.05	0.46	-2.30	0.022
lingual1	0.22	0.17	1.26	0.206
condition1:ot1	-2.77	0.83	-3.35	<0.001
condition1:ot2	0.90	0.44	2.02	0.043
condition1:lingual1	-0.17	0.17	-1.00	0.315
ot1:lingual1	0.60	0.82	0.73	0.467
ot2:lingual1	-0.39	0.43	-0.91	0.365
condition1:ot1:lingual1	-0.91	0.81	-1.12	0.264
condition1:ot2:lingual1	0.09	0.43	0.20	0.843

**Table 3:** Results from Experiment 5 with Spanish native speakers, CLASS accuracy in the subset of the data with incorrect responses. The predictors of condition and number of L1s were effects-coded with agglutinating and monolingual being compared to the grand average.

	Estimate	Std. Error	z value	Pr(>  z  )
(Intercept)	1.08	0.13	8.39	<0.001
condition1	0.11	0.13	0.85	0.394
ot1	8.65	0.62	13.88	<0.001
ot2	-1.60	0.30	-5.39	<0.001
L11	-0.11	0.19	-0.60	0.546
L12	-0.24	0.19	-1.23	0.218
condition1:ot1	0.06	0.60	0.10	0.918
condition1:ot2	-0.26	0.27	-0.97	0.334
condition1:L11	0.09	0.19	0.47	0.635
condition1:L12	0.04	0.19	0.20	0.844
ot1:L11	-0.70	0.89	-0.79	0.428
ot1:L12	-0.09	0.90	-0.10	0.923
ot2:L11	0.49	0.40	1.20	0.228
ot2:L12	0.19	0.41	0.46	0.645
condition1:ot1:L11	0.22	0.88	0.25	0.799
condition1:ot1:L12	-0.05	0.89	-0.06	0.955
condition1:ot2:L11	-0.14	0.40	-0.34	0.734
condition1:ot2:L12	0.15	0.41	0.38	0.706

**Table 4:** Results from the model on the combined data of all monolingual participants in Experiments 1 (English native speakers), 4 (Turkish native speakers) and 5 (Spanish native speakers), predicting overall accuracy. The predictors of condition and L1 were effects-coded with agglutinating, Turkish and Spanish being compared to the grand average.

	Estimate	Std. Error	z value	Pr(>  z  )
(Intercept)	0.62	0.10	6.07	<0.001
condition1	-0.38	0.10	-3.74	<0.001
ot1	4.55	0.51	8.85	<0.001
ot2	-1.21	0.32	-3.76	<0.001
L11	-0.07	0.15	-0.48	0.632
L12	0.23	0.15	1.47	0.141
condition1:ot1	-1.74	0.51	-3.41	<0.001
condition1:ot2	0.15	0.32	0.49	0.626
condition1:L11	0.01	0.15	0.07	0.941
condition1:L12	-0.31	0.15	-2.02	0.043
ot1:L11	-0.59	0.76	-0.77	0.439
ot1:L12	1.22	0.76	1.60	0.109
ot2:L11	0.55	0.48	1.14	0.253
ot2:L12	-0.51	0.47	-1.08	0.278
condition1:ot1:L11	0.18	0.76	0.23	0.817
condition1:ot1:L12	-1.52	0.76	-2.00	0.045
condition1:ot2:L11	-0.17	0.48	-0.35	0.723
condition1:ot2:L12	0.38	0.47	0.81	0.419

**Table 5:** Results from the model on the combined data of all monolingual participants in Experiments 1 (English native speakers), 4 (Turkish native speakers) and 5 (Spanish native speakers). The model predicts CLASS accuracy in the subset of the data with incorrect responses. The predictors of condition and L1 were effects-coded with agglutinating, Turkish and Spanish being compared to the grand average.

## Chapter 6

### General Discussion

#### 6.1 Aims of this thesis

The impact of transparency has been investigated over many decades in language acquisition research, from Brown (1976) to Goldschneider & DeKeyser (2001) to Narasimhan & Gullberg (2011) and many others. Early on, during the 1970s and 1980s, but also more recently in the study by Goldschneider & DeKeyser (2001), transparency was often termed ‘semantic complexity’. In their quest to find factors that are responsible for acquisition ease and difficulty with regards to morphology, morpheme order studies proposed and tested various factors pertaining to language-internal as well as language-external factors. Semantic complexity was one of them and the effect that it was found to have on morpheme acquisition varied (Andersen, 1978; Brown, 1976; de Villiers & de Villiers, 1973; Larsen-Freeman, 1976). However, linguistic descriptions often invoke transparency in terms of one-to-one mappings between meaning and form as an ideal feature of language (e.g., Bybee, 1985; Carstairs-McCarthy, 1987; Don, 2017; Kusters, 2003). More recent studies have again investigated the effect of transparency on learnability and drawn a more consistent conclusion, namely that transparency is an important factor determining acquisition ease (Bittner et al., 2003; Korecky-Kröll & Dressler, 2009; Narasimhan & Gullberg, 2011). On a more general level, it has also been proposed that agglutinating systems are easier to learn than fusional systems, again based on a difference in transparency (Aksu-Koç & Slobin, 1985; Bittner et al., 2003; Hengeveld & Leufkens, 2018; Igartua, 2015). However, these studies have one main commonality: they trace the effect of transparency in studies of natural languages, where morphological paradigms have many inter-related characteristics.

In their meta-analysis of studies on the acquisition of English grammatical morphemes, Goldschneider & DeKeyser (2001) find that five predictors (semantic complexity, perceptual salience, morphophonological regularity, syntactic category and frequency) combined can explain over 70 per cent of the variance in acquisition order. However, the variables are highly correlated in their model, which is why the authors propose to “try to tease apart the individual determinants’ effects on acquisition in order to establish with more certainty whether the combination of factors

accounts for the order through a cumulative effect or through the interaction of the factors” (Goldschneider & DeKeyser, 2001, p.38). The strength of the artificial language paradigm I used in this thesis is that it allowed me to do exactly this and to probe the role of transparency in a relatively isolated manner. A factor that was not included in the study by Goldschneider & DeKeyser (2001) is L1 transfer, which has also often been suggested and sometimes tested in the wake of the morpheme order studies (see Jarvis, 2015). This thesis combines an investigation concerning the role of transparency in the acquisition of inflectional morphology with the influence of morphological experience from the mother tongue. Specifically, the experiments reported here test a potential link between learnability and morphological techniques – isolating, agglutinating, fusional – in the L1 and L2.

Importantly, the method used in this thesis made it possible to exclude and reduce the influence of additional factors that typically correlate with transparency in morphological paradigms in natural languages. The number of grammatical morphemes to be learned in each condition was held constant at four; this is the only size of meaning space for which the required number of morphemes is the same between a compositional (e.g., agglutinating) and a holistic (e.g., fusional) system. Further, the high variability of word stems in the experiment, along with adhering to the phonotactics of each respective L1, directed participants’ attention to the grammatical morphemes, which were fully regular. Two other factors typically play a role in the morphology of natural languages. Agglutinating morphology is largely compositional, in that the overall meaning of a word can be derived from the meanings of the individual concatenated morphemes (Dressler et al., 1987; Vincent, 2014). Fusional morphology, on the other hand, is holistic, since individual fusional morphemes typically express multiple meanings which cannot be separated onto individual forms. Due to this compositional nature, agglutinating systems are generalisable and morphemes occur more frequently than those of holistic systems. However, the generalisation advantage of the agglutinating condition was relatively small, since four morphemes were repeated across 96 trials. The agglutinating morphemes occurred twice as often as the fusional morphemes, but again considering the low number of distinct morphemes across trials, their frequency was high in both conditions.

Overall, the chosen design therefore ensured that transparency was the most important contrast between the two experimental conditions. These corresponded to

an agglutinating and a fusional morphological paradigm consisting of two features, NOUN CLASS and NUMBER. The level of morpheme transparency was modulated by manipulating exponence, which is considered a central feature in the delineation of morphological types (Bybee, 1985; Haspelmath, 2009; Plungian, 2008). The two features were mapped onto morphemes either in a transparent way (a one-to-one mapping between form and meaning) or in a non-transparent way (a one-to-many mapping between form and meaning). The impact of transparency on learnability was tested with three different groups of learners representing three different morphological types, isolating (English), agglutinating (Turkish) and fusional (Spanish). This way, it was possible to investigate whether the morphology of a learner's L1 affects the proposed benefit that transparency is claimed to hold for learnability.

## 6.2 Summary of predictions

Considering the experimental design and the participant groups, several predictions were derived based on previous literature prior to conducting the experiments. Regarding native speakers of English, who are less experienced with inflectional morphology through their L1 and therefore relatively neutral to the agglutination-fusion distinction, I predicted that transparent forms would be easier to learn as previously claimed, which should be reflected in higher overall accuracy rates in the agglutinating condition. This was the main prediction being investigated in all experiments presented in **Chapter 3**. In **Chapters 4** and **5**, two further experiments were devised so that in total, the influence of different types of morphological structures in the L1 on the acquisition of L2 morphology could be investigated.

Research suggests that L1 transfer can happen at different levels of the language, including transfer of morphological processing strategies and transfer of abstract features. Research has found that morphological structures in the L1 can impact learning of L2 morphology in different ways. A study by Havas et al. (2015) found that experience with morphological processing was more beneficial for the acquisition of an abstract grammatical feature than experience with the feature through the L1, a result which they tentatively attribute to better detection and awareness of morphological forms due to participants' morphologically richer L1. In turn, this could have helped participants learn the corresponding meaning of the

morphemic units. It has also been indirectly suggested that the morphological technique in the L1 could lead to a sort of abstract category transfer, i.e., to a learning preference for the morphological technique in the L2. For example, Igartua (2015) suggests that the emergence of agglutinating features in otherwise fusional morphological paradigms in various languages is due to language contact situations where L1 speakers of largely agglutinating languages prefer this morphological technique in the L2 (fusional) morphology. While research on this sort of transfer is sparse, more studies have been conducted on the role of the L1 for L2 morphological processing.

More precisely, it has been found that L2 learners with inflectionally very rich L1s use morphological decomposition to a larger extent in their L1 compared to L1 speakers of languages that are inflectionally poorer, and that decomposition is also used more often by these learners in the L2 (Portin et al., 2007; Portin et al., 2008). While studies have rarely investigated this effect in beginning L2 learners, it has been shown that such processing effects exist at different levels of proficiency and that L2 speakers show L1-like morphosyntactic sensitivity in offline tasks while online processing differences still exist (Bosch et al., 2017; Gor & Jackson., 2013; Kimppa et al., 2019). Finally, studies suggest that when morphologically complex words are observed with lower frequency (as in bilingual speakers), this leads to increased decomposition (Lehtonen & Laine, 2003; Lehtonen et al., 2006). While this thesis did not test the type of processing employed by participants, it allows for a comparison of the outcome, namely acquisition success, in relation to different L1s. Thus, in line with research showing that morphological processing is more prevalent than full-form access in L2 learners and in less frequent words, I predicted that native speakers of Turkish and Spanish, both inflectionally richer languages than English, will show facilitative L1 learning effects in the present experiments (Gor & Jackson, 2013; Portin et al., 2007). However, I also predicted that their experience with different types of inflectional morphology through the L1 will differentially affect learning in the experiment.

Based on the assumption that morphological decomposition will be transferred from the Turkish L1 and that decomposition is helpful in the detection of morphemes and their grammatical meanings in an L2, I predicted that the group of Turkish native speakers would show higher overall accuracy rates than native speakers of English in both conditions. I also expected that they would show the largest bias for transparent

structures out of all three L1 groups and perform significantly better in the agglutinating than in the fusional condition. This prediction is additionally supported by the fact that Turkish is not only rich in inflections, but also highly agglutinating. Thus, the morphological structure prevalent in the L1 and the structure that is generally predicted to be easier to learn coincide for this participant group. I further predicted a transparency bias borne out in terms of a stronger focus on CLASS than NUMBER in the fusional condition, since this strategy would render fusional morphemes transparent.

Recall that NUMBER was unambiguously binary in the experiments (one versus two entities in the images). On the other hand, the ANIMACY sub-classes of animate and inanimate were less straightforward to discern. Since ANIMACY was an intrinsic part of the images, they could have been categorised based on many different criteria, such as size, land and sea animals or tools and clothing. Crucially, the more open hypothesis space within ANIMACY means that it was possible to find four sub-classes of ANIMACY instead of two. This was not possible for the binary NUMBER category. What this means in turn is that participants could ignore NUMBER and look for four sub-classes based on ANIMACY, which could be mapped onto the four different suffixes they saw on the screen. This interpretation overlaps with participants' responses in the questionnaire.

Observing that participants learned CLASS better than NUMBER (that is, participants producing more responses with correct CLASS compared to correct NUMBER) is an indicator that participants focused more on CLASS. This led to errors such as selecting an animate singular response in an animate plural trial: if a participant makes a distinction between land and sea animals and ignores NUMBER, the selected answer still correctly indicates CLASS, since both are animate, but may indicate the wrong NUMBER. Under the hypothesis that the four grammatical suffixes indicate only CLASS distinctions, fusional morphemes are treated as transparent, since participants assume that they only express one grammatical feature, namely ANIMACY. If a bias for transparency exists, CLASS correct only responses should be more frequent in the fusional condition, since agglutinating morphemes are already transparent and should allow for the acquisition of both features. Assuming amplification of this error type through the transparent L1, I supposed the proportion of CLASS correct only responses to be even higher for L1 Turkish speakers than it was for L1 English speakers. This assumption is strengthened by the fact that knowledge of abstract features in the L1 can exercise positive transfer effects (Sabourin et al., 2006). Since

Turkish requires some sensitivity to ANIMACY distinctions, this should have helped the L1 Turkish participants learn this feature in the artificial language.

The group of L1 Spanish speakers was also expected to benefit from their larger experience with inflectional morphology compared to native speakers of English and to perform better in the experiment overall than native speakers of English. However, since Spanish is largely fusional, these speakers' experience with the particular type of morphology in the L1 clashes with the suggested benefit of transparent morphology. Therefore, their bias for transparent structures might be smaller compared to the L1 Turkish group. Nevertheless, given the claims in the literature, I expected this group to also learn better in the agglutinating than in the fusional condition. Both for native speakers of Turkish and Spanish, it was predicted that the advantage of agglutinating systems would be borne out due to their inflectionally richer L1s, which I considered a hindering factor in the L1 English group. While GENDER is absent in Turkish, it exists in Spanish, and Spanish also shows ANIMACY effects in the form of Differential Object Marking. Since both ANIMACY and GENDER are examples of NOUN CLASS, L1 Spanish speakers were also predicted to show a stronger focus on CLASS compared to the L1 English group, that is, the proportion of CLASS correct only responses was expected to be larger in the L1 Spanish compared to the L1 English group at least in the early phase of the experiment. Performance of all three L1 groups was compared in a final statistical model.

To summarise, given the strong claim in the literature concerning a learning benefit for transparent structures, I generally predicted the agglutinating system to be learned better and faster than the fusional system in all three L1 groups. However, I expected this advantage to be exploited more by the group of L1 Spanish speakers and especially by native speakers of Turkish. This general hypothesis of better learning in the agglutinating condition is even slightly strengthened by the two small unavoidable confounds inbuilt into the artificial language, namely the earlier generalisability of the agglutinating system and the higher frequency of the agglutinating morphemes. Since generalisability and frequency have both been shown to be helpful for language learning, if these small differences did have any effect in the first place, they should have increased the learnability of the agglutinating condition.

### 6.3 Findings contradicting the claim surrounding the impact of transparency

Contrary to the main prediction made here, the findings in this thesis are unequivocal in that none of the analyses of overall acquisition showed a significant effect of transparency, the only exception being Experiment 3 which tested length-matched fusional morphemes without colour. As discussed in **Chapter 3**, however, this effect appears to be driven by segmentation difficulty in the fusional non-colour condition, not by an advantage of the agglutinating system. Therefore, the findings of the experiments presented in this thesis stand in sharp contrast to the frequent claim that transparent morphemes are easier to acquire than non-transparent ones (Aksu-Koç & Slobin, 1985; Bittner et al., 2003; Hengeveld & Leufkens, 2018; Igartua, 2015; Penke, 2012). Across three different speaker populations, I found no learnability advantage of the agglutinating system. After native speakers of English did not show a preference for transparent structures in Experiment 1, I additionally tested two potential confounds, number of morphemes per word and word length, both of which are greater in agglutinating systems. This was done to see whether their effect was masking a learnability benefit in Experiment 1. However, the results from Experiments 2 and 3 in **Chapter 3** showed that neither segmentation effort nor longer words negatively affected learning in the agglutinating condition. Results from Experiment 2 suggest that participants were segmenting the two agglutinating morphemes well even in the absence of colour cues. However, since English is inflectionally poor, segmentation might have been costly for these participants and two further groups of participants were tested whose L1s are inflectionally richer. However, neither native speakers of Spanish, nor native speakers of Turkish, one of the near textbook examples of agglutinating systems, learned better in the agglutinating compared to the fusional condition. These individual findings were confirmed in a model of the combined data from Experiments 1, 4 and 5, which again revealed no effect of condition and no learnability differences between L1 speaker groups.

The pattern of results looks somewhat different when considering the error patterns rather than overall accuracy. Results are again very similar across experiments, with two important points to note. First, participants produced significantly more CLASS correct only responses than other error types in both conditions across all five experiments. Second, the proportion of CLASS correct only

responses among incorrect responses was higher in the fusional than in the agglutinating condition, which is confirmed by a model of the combined data from Experiments 1, 4 and 5. This observation can be explained by a strategy whereby participants are searching for four different values within the CLASS feature, such as land and sea animals versus household items and clothing. Strategies along these lines, where the values of animate and inanimate were split up further, were indeed frequently reported in the final questionnaires in each of the five experiments. By focusing entirely on CLASS and ignoring NUMBER, participants are finding a single feature with four values, which renders the fusional morphemes transparent. This bias for transparent morphemes was amplified by colour cues; when the single fusional morpheme was highlighted, participants were more likely to treat it as transparent. The difference in the proportion of CLASS correct only responses between conditions was largest for the group of Spanish native speakers, which could be interpreted in two ways. Either it is due to a bias for transparent structures or due to the fact that ANIMACY and GENDER, both instances of NOUN CLASS, are familiar to native speakers of Spanish. It was regarded more plausible that this pattern is due to the slight bias for transparent morphemes, which better explains the discrepancy between conditions. In any case, it does not suggest that L1 Spanish speakers have a stronger preference for fusional morphology. Overall, the pattern of errors from all experiments suggest a bias for transparent structures since the effect replicates across experiments. This suggests that a transparency bias might exist, and in the experiments reported here, it is strong enough to lead participants to ignore the NUMBER cue, which was salient in the images. On the other hand, the bias is too weak to be reflected in higher overall accuracy in the agglutinating condition. The experimental findings are therefore consistent in two ways: there appears to be some evidence for a transparency bias, but it is much weaker than commonly stated, when transparency is tested in isolation.

One might be tempted to argue that the findings pertain in large part to the particular experimental design used in this thesis. However, ANIMACY and NUMBER are not only easy to depict in images, they also occur frequently as inflectional features in natural languages and are therefore neither rare phenomena nor cognitively highly demanding. Research also suggests that humans are sensitive to ANIMACY distinctions (Blumenthal et al., 2018; Bonin et al., 2019; Kriegeskorte et al., 2008; Calvillo & Jackson, 2014; Strickland et al., 2017). Nevertheless, it appears a valid question to ask whether the results might have been different had the salient feature of ANIMACY

been replaced by a different one. Based on participants' responses in the questionnaire, both conditions featured participants who only learned CLASS, but not NUMBER, or who learned CLASS before NUMBER. Even if this issue was more pronounced in the fusional condition, this is essentially a problem of learning grammatical categories, not their morphemic correspondences. Replacing CLASS by a slightly less salient feature could have therefore meant that both features would be learned earlier and produced higher overall accuracy scores, but this does not straightforwardly map onto faster learning of either transparent or non-transparent morphemes. Further, learners have been found to show more acquisition difficulty with longer words (Laufer, 1990), but the experimental words consisted of monosyllabic stems followed by one or two suffixes, which is about the minimum required to be able to contrast transparency levels of inflectional morphemes. Finally, effects of novel phonotactics were avoided in the experimental design and transparency was the main difference between conditions. If transparency does have a considerable impact on learnability, there is strong reason to expect it to show in at least some of the five experiments that were conducted. Any claims that the lack of the expected advantage is due to experimental design therefore need to be formulated in relation to a precise reason.

#### **6.4 Potential limitations of the experimental design**

It is worth considering a few design features of the experimental task which might have impacted the results. First, the experimental session in the experiments reported here was short, as in most artificial language learning studies (about 25 minutes). Of course, morpheme acquisition studies track learning over long periods of time or use cross-sectional approaches (Bittner et al., 2003; Brown, 1976; Korecky-Kröll & Dressler, 2009; Kwon, 2005). One might therefore argue that the fact that a transparency advantage is not borne out in the experiments used in this thesis is due to the short duration of testing. However, it does not seem obvious how a longer testing time would impact the role of transparency on learnability – if the morphological paradigm was larger, containing more morphemes, it appears logical that learning the forms would require more time, and agglutinating compositional structures might be difficult to acquire at first. However, the two-by-two paradigm used here only contained four morphemes in each condition and they were learned relatively quickly by at least

some of the participants. For this reason, it is difficult to see how longer exposure time would have led to a more prominent advantage of transparency – more likely, it would have led to ceiling effects for all learners in all conditions. The experiments reported here appear to have achieved a good balance: the task is doable, but not trivial and overall accuracy levels are around 60 per cent, with performance around 75 per cent on average toward the end of the session.

Another issue is the retrieval of form-meaning mappings in the agglutinating condition, which may have been more difficult than in the fusional condition, since two rather than one morpheme was involved per word. This also meant that morphemes in the agglutinating condition occurred in a fixed order, something which technically needs to be learned as well. However, the task employed did not require participants to produce words and due to the small size of the paradigm, this is not expected to impact the results considerably. It has also been found that acquiring the order of morphemes is not very difficult (Leonard, 2014; Slobin, 1973).

Another point to mention is that participants across experiments were tested on two different platforms, Amazon Mechanical Turk and Prolific. In principle, it is possible that participants will show different behaviour depending on the platform, but there do not seem to be any differences in performance when considering the two final statistical models presented in **Chapter 5**. Native speakers of Turkish, tested on Prolific, showed very similar learning patterns to native speakers of Spanish and English, who were tested on Amazon Mechanical Turk. If Prolific had a negative effect on acquisition success, which masked the transparency advantage, it is difficult to explain why learning in native speakers of Turkish was not different from that of Spanish native speakers, who are also familiar with inflectional morphology. It is also difficult to see why differences in learning patterns were not found between L1 English and L1 Spanish speakers, who differ in terms of morphological experience, but were tested on the same platform. Moreover, responses in the questionnaire were similar between all three groups in terms of learning strategies and areas of difficulty that participants identified.

## **6.5 Implications of the experimental results**

Having excluded several language-external and -internal factors as potential reasons for the lack of a transparency advantage, we can now turn to the implications of the

experimental findings for language learning and for acquisition research. What follows from them is that a particular morpheme is not necessarily learned better than another morpheme expressing the same meaning, only on the basis of it being more transparent. When trying to find direct analogues between the experimental paradigm tested here and natural languages, we would expect that there is no learnability difference between a natural language paradigm expressing two binary features (such as binary NUMBER and binary CASE) using separative exponence and a paradigm which expresses these features through cumulative exponence. Another example might be binary NUMBER and binary PERSON or TENSE in verbal inflection. If learnability differences exist, they should be due to other characteristics of the system. Put differently, when the morphological paradigms in question are small, transparency differences are not expected to lead to a learning advantage of the more transparent system. However, this does not mean that transparency does not have an impact under *any* conditions, and there is reason to believe that the advantage takes effect when the systems become large (Brighton, 2002; Plank, 1986). In language acquisition research, it has been shown that there are learnability differences between larger morphological systems, such as verbal morphology in Turkish compared to Dutch in children (Aksu-Koç and Slobin, 1985; De Houwer & Gillis, 1998; Hengeveld & Leufkens, 2018). In larger systems, however, other factors likely play a role as well, such as regularity, number of morphemes in the system and morpheme frequency. Therefore, any learnability effects that are found might not be due to transparency alone, but to its interaction with other factors. This interpretation is supported by the fact that we see a bias for transparency in error patterns, suggesting it exists, but is relatively weak in isolation.

An important precondition for the acquisition of inflectional morphology is that the morphemes in question are noticed by the learner (Penke, 2012). Since native speakers of English are not very familiar with inflectional morphology through their L1, segmentation of the morphemes was initially considered a potential cause of difficulty. However, in Experiment 2, where a facilitation effect of visual cues to morpheme boundaries was not found, there was evidence that at least offline segmentation was happening successfully, including in the agglutinating condition. This is an important point to raise, since claims about the benefit of transparent structures indirectly assume that learners treat transparent morphemes as individual units. In order to test whether segmentation was nevertheless costly for native speakers of English, two

further participant groups were tested. In line with Luk & Shirai (2009), this thesis thus attempted a differentiated look at the learning of L2 morphology by speakers of different native languages and by investigating potential transfer effects of L1 morphological structures onto the L2. Although Luk & Shirai (2009) suggest that L2 acquisition can follow very different pathways depending on the L1 system of a learner, I did not find any considerable differences between L1 speakers of English, Turkish and Spanish in the acquisition of transparent versus non-transparent morphology.

This can suggest two things. On the one hand, facilitation through the morphological structure of the L1 may exist, but may not be happening in the task presented here. On the other hand, there may be facilitation through the L1 also in the earliest stages of L2 acquisition, as tested here, but they do not lead to better learning in the transparent condition, since its benefit is small. Given that the experiments presented here do not tap into the particular kinds of offline and online processing that are taking place, it is difficult to tell which of the two alternatives plays a larger role in this experimental setup. Many studies on processing effects resulting from L1 structures have been conducted with learners who were no longer absolute beginners (Gor & Jackson, 2013; Kimppa et al., 2019; Portin et al., 2008). It is therefore possible that a certain level of L2 proficiency is required before processing techniques are transferred onto the L2. On the other hand, Havas et al. (2015) found an effect of morphological richness in the L1 in an artificial language learning paradigm. If Turkish native speakers were indeed decomposing the agglutinating morphemes more readily than native speakers of English and also Spanish, we would assume that this led to greater awareness of the morphemes, making it easier to learn their underlying meanings. However, if the advantage of transparency, which was the largest difference between the conditions, is small, then this will not enhance learning significantly even with greater ease of morphological processing. Moreover, neither transparency nor lack of transparency in morphemes in the L1 appears to influence the results, which suggests that there is no strong bias for transparent structures, or morphological structures to be similar to the mother tongue, at least not in the context of a relatively small system. Thus, knowledge of certain morphological structures is not always a guarantee for better learning, but depends on the system that is being learned. What particular types of processing are at play in the present tasks would be a topic for future research.

## **6.6 The problematic focus on transparency in acquisition research based on linguistic theory**

Having described the experimental results of this thesis and their implications for acquisition research, let us return to transparency and ask why a carefully designed large-scale experiment does not produce evidence for the widely assumed transparency benefit. In this section, I will take a closer look at the studies of natural languages I referred to earlier and the kinds of evidence that these claims are based on. I will also review some of the statements made about transparency in linguistic theory, which some of the research on natural languages refers to. First, let us return to the studies I described in **Chapter 2 (Sections 2.1.4 and 2.1.5)**, which motivated the current research. This will show how they all have a certain commonality, namely that transparency was not strictly isolated from other factors. Examples of this pattern will strengthen the inference that transparency impacts on learning ease by means of an interplay between multiple factors.

Early on in morpheme order research, Brown (1976) and de Villiers & de Villiers (1973) pointed out that in their analysis, semantic complexity could not be fully distinguished from grammatical complexity, making it impossible to tell which of the factors was driving the acquisition patterns they described. They endorsed transparency as a factor of interest, but testing its impact remained difficult. Another prime example of correlated factors in natural languages is the meta-analysis by Goldschneider & DeKeyser (2001). They coded the English grammatical morphemes for five different predictors (semantic complexity, frequency, morphophonological regularity, perceptual salience and syntactic category). However, as they point out themselves, the size of the effect of each individual predictor cannot be determined due to the amount of correlation that exists between the predictors in their model. While the findings from this study are highly important in showing that the different factors affect learning in combination, they cannot support the claim that transparency plays a major role.

Correlation of factors also plays a role in the study by Laaha & Gillis (2007) in which they predict the development of nominal and verbal morphology in children by the factors of paradigmatic and syntagmatic morphological richness, morphological transparency, uniformity and salience. In their description, they name transparency as the second most important predictor after morphological richness, highlighting its

conceptual significance. However, in their analysis of longitudinal production data, they find that the effect of transparency is no longer significant when controlling for paradigmatic morphological richness (the number of inflectional categories that nouns and verbs inflect for), which suggests that the latter is the underlying predictor of speed of morphological learning. These findings again suggest that transparency of individual forms can be correlated with other factors in natural languages, in this case with morphological richness. This is strengthened by the fact that Laaha & Gillis (2007) tested the effects with data from typologically different languages. In line with one of the conclusions drawn in this thesis, the authors note that effects might be different if the determinants were tested in combination rather than in isolation.

In their research on the development of small verbal paradigms (termed 'miniparadigms') across different languages in children, Bittner et al. (2003) find that the interval between children's production of first verbs and their first miniparadigms differs depending on the morphological character of the language. The interval is smaller for morphologically richer and more transparent languages in their sample, such as Turkish, Finnish, Russian and Croatian. Bittner et al. (2003) relate this to factors like productivity and transparency, while at the same time, they note that a single language displays many characteristics. For instance, they describe that Turkish mostly lacks allomorphy, fusion and inflectional classes, which means that it is very regular and contains relatively few morphemes per inflectional category, which by implication repeat rather frequently. How much each individual aspect of the morphological system in question contributes to learning ease is not tested in their study. This example therefore again highlights how a combination of multiple factors may predict acquisition differences between languages, but the exact role of transparency in this process is difficult to determine.

Korecky-Kröll & Dressler (2009) provide a detailed description of how eight plural markers in German compare with regards to productivity, transparency, iconicity as well as type and token frequency, from which they are able to derive predictions as to which plural markers should be acquired earlier than others. This is a systematic and careful approach to test the importance of characteristics of the morphological forms. For instance, -s plural (as in *Auto-Autos* (car-cars)) is predicted to be acquired earlier than -e + *umlaut* plural (as in *Bank-Bänke* (bank-banks)): the former is productive, transparent and iconic, but infrequent, while the latter is only partly productive, less transparent due to the umlaut and of medium frequency. While the

differences between the levels for each parameter are well-described, the categorisation of markers as either transparent or non-transparent, etc., appears to some degree arbitrary. Further, while markers can be directly contrasted which only differ in one parameter, it is difficult to quantify its importance, even if one marker is learned earlier than the other one. At the same time, since markers display various parameters, one may again expect some correlation between the different factors. This illustrates how even carefully planned studies of natural languages run into difficulty testing predictors in isolation.

Narasimhan & Gullberg (2011) compare the influence of frequency and transparency in caused posture verbs in Tamil and Dutch. In Tamil, these verbs are transparent, but infrequent, while in Dutch, they are non-transparent, but frequent. Their study finds that transparency has a larger impact on learnability than frequency, since caused posture verbs are mastered by the age of four by Tamil-speaking children, but still in the process of acquisition at age five by Dutch-speaking children. Their study targets acquisition of a construction which appears very comparable between the two languages, although not all language-internal factors can be controlled for. The context in which caused posture verbs occur in both languages may exert influence beyond the frequency of forms, since contextual factors appear to be more complex in Dutch than in Tamil. Although the authors note that the conceptual distinctions relevant to the use of the verbs are learned well by the children in their study, having to learn more distinctions may still impact negatively on acquisition. However, it is interesting to observe that Dutch-learning children produced forms which were ungrammatical, but transparent, which in some way parallels the error patterns observed throughout this thesis. Again, there may indeed be a bias for transparency of forms, but testing it fully independently of other factors remains difficult.

As these examples from studies on natural languages illustrate, there is a clear similarity in that transparency in these languages is always correlated with other characteristics of the inflectional systems in question. While it is possible to carefully describe these characteristics and to compare markers with otherwise little differences, transparency can never be tested fully in isolation. Against the backdrop of the studies above, claims made by Hengeveld & Leufkens (2018) appear too simplistic. A statement from their study illustrate this. Assuming that natural languages

can be ranked in terms of their transparency as the authors do in a very comprehensive way, they further elaborate that

“it has been claimed that children acquire transparent features of languages much faster than non-transparent features, so if some languages can be said to be more transparent than others, we have strong indications that, barring other factors, these languages are easier to learn as well” (Hengeveld & Leufkens, 2018, p.140).

The problematic aspect of this statement is that claims about the role of transparency for acquisition are adapted from earlier literature, without discussing them in much detail, but they are used to formulate a rather strong prediction. This is reflected in another statement:

“we argue that transparency is an important factor contributing to the (L1) learnability of languages, as a result of which the placement of languages along the transparency hierarchy can be interpreted in terms of the ease or difficulty with which they will be acquired” (Hengeveld & Leufkens, 2018, p.171).

Again, the claim about transparency affecting acquisition is not tested in their study, but used to formulate a new prediction (concerning entire languages). Given the findings from this thesis, we should use some caution to not overestimate the influence of transparency, especially since it appears that transparency interacts with other factors, such as regularity, frequency, perceptual salience, etc. Although Hengeveld & Leufkens (2018) consider transparency beyond the morphological level, this does not preclude that it also interacts with factors elsewhere.

Do these difficulties with studies on natural languages mean that the role of transparency has been overrated? The results presented in this thesis suggest so. When looking at developments towards more transparency across languages, as in research by Anttila (1989) and Igartua (2015), we can see use of the argument that transparency is more ideal or natural, but a concrete link between transparency and learning ease is not provided, as I will illustrate below. Instead, what we see for instance with Igartua (2015) is that the theory-based Transparency Principle as presented in Kusters (2003) is taken over into the description of natural languages. The remainder of this section will address the importance of making a distinction between linguistic theory and empirical research, which is concerned with causal links in the data. The acceptance of claims surrounding transparency without much critical review is not a symptom of this particular study alone; transparency seems to enjoy high reputation among linguists. Besides research on the role of transparency in

language acquisition, studies have also related transparency to language change. Here, problematic issues do not arise so much from the intricacies of natural languages, but from the lack of an empirical test of the causal relationship between transparency and learning ease. In this way, the studies are similar to Hengeveld & Leufkens (2018), who do not discuss the reason for a transparency bias either.

In a chapter on analogy across languages, Anttila (1989) remarks that “[l]anguage has a general iconic tendency, whereby semantic sameness is reflected also by formal sameness” (p.89). Anttila (1989) bases this claim on his observations which show that analogy sets in and restores regularity in systems which had previously become irregular due to sound changes. However, based on the dichotomy between sound change and analogy that he describes, one could also argue that language has a tendency towards irregularity. The sense of an ascription of a preference emanating from language itself also appears in another quote from his book where he says that: “the iconic basis of language, a preference for parallelism (one-to-one relation) between meaning and form, prevailed” (Anttila, 1989, p.98). In the text excerpt below, Anttila (1989) highlights how the notion of a tendency in language to be transparent has a long history in theoretical descriptions of language. This is even more compelling in light of the fact that Anttila (1989) notes that language is both transparent and non-transparent. Where this tendency in language towards ‘one meaning one form’ originates from, is not clear, it remains a theoretical description:

“The higher, more general principle of ‘one meaning, one form’ is as old as European linguistics. It has been referred to, among other things, as the principle of optimality (Humboldt), or univocability (Vendryes), and as the canon of singularity (Ogden and Richards). [...] It has always been known that this principle is a tendency only, like so much in human behavior and biology that is not susceptible to rigorous formulation. No one has ever implied that it actually would lead language to a point where every meaning would have its own form, or total one-to-one correlation between form and meaning” (Anttila, 1989, p.107).

A similar resort to the theoretical ideal of transparency can be found in Igartua (2015) who claims that changes from fusion to agglutination (which constitute a reversal of the morphological cycle) in a sample of languages are likely related to language-external factors stemming from language contact. He largely links his claims to statements from other authors (Dahl, 2004; Kusters, 2003; Miestamo, 2008), noting

that transparent structures are simpler, easier to acquire for adult learners, or that some agglutinating changes in his sample led to language simplification, which learners in turn prefer. However, how transparency of form-meaning mappings relates to learning ease more specifically is not discussed in any detail in this study. Simplifications also occur through the regularity of using an invariant plural marker throughout, for instance in Modern Eastern Armenian which he studies in his sample, which leads to regularity as well as to transparency of the marker in question.

Igartua (2015) does acknowledge those simplifications that lead to other characteristics, such as regularity and lack of allomorphy. Fewer markers and fewer rules to remember additionally increase the simplicity of a system. Since these are all related to agglutinating systems, as is transparency, what is missing is a clear demonstration – theoretical or empirical – in how transparency specifically leads to learning ease. Correlated factors certainly complicate the picture and make it difficult to test the proposed link empirically, but Igartua (2015) does not offer a differentiated description of this link. It may certainly be the case that in the language-contact situations he studies, agglutinating languages exert considerable influence on the neighbouring fusional languages. Whether this is primarily caused by transparency, however, is questionable. It is possible that agglutination is preferred over fusion by adult learners for a number of reasons such as its regularity, reduced number of markers, compositional structure, etc., but given the empirical findings presented in this thesis, the role of transparency may be overstated. The nominal and verbal paradigms that Igartua (2015) studies are also all larger than the two-by-two paradigm used for this thesis, making it highly likely that transparency is not the only factor at play. In this vein, statements like morphological changes being “guided by the transparency principle” (Igartua, 2015, p.712) may need to be reconsidered. Changes in languages may lead to more transparent patterns, but whether transparency is the cause of this change requires further investigation.

While claims about a positive impact of transparency are widespread, studies often appear to be motivated by a theoretical notion of transparency. For instance, Igartua (2015) concludes that “the transparency principle leads to the avoidance of (...) ‘learner-unfriendly phenomena’” (p.709). Kusters (2003) describes the Transparency Principle as dating back to research by Wilhelm von Humboldt on morphological typology. Kusters (2003) adds that “[t]he Transparency Principle demands that the relation between form and meaning is as transparent as possible”

and further that “[t]ransparency decreases when words deviate from this ideal” (p.21). What is important, however, is that the Transparency Principle is a theoretical concept. Therefore, when trying to explain linguistic phenomena based on transparency, the underlying mechanism (or causal relationship) needs to become explicit. A trend towards transparency can certainly be a fitting description from a linguistic point of view, but whether transparency is also the driver of change is a separate question.

Carstairs-McCarthy (1987) invokes transparency to describe a possible state of language, namely one in which one exponent expresses one grammatical meaning, against which actual states of language can be compared. That is, by describing a transparent form-meaning mapping, it becomes possible to observe instances which deviate from this ‘ideal’ transparent mapping. The study by Korecky-Kröll & Dressler (2009) also tests the importance of transparency, productivity and iconicity, which are principles in Natural Morphology. Here again, these principles are a theoretical description of language, which the authors go on to test in the production of a child acquiring Austrian German. Transparency of forms is not a given preference of a child learner. As I illustrated above, it is indeed rather difficult to demonstrate how transparency impacts on learnability.

In Natural Morphology, the one-to-one principle describes that the mapping between meaning and form is uniform and transparent, that is, bidirectional, again ascribing a “special status” to this particular type of mapping (Carstairs-McCarthy, 1987, p.13). Carstairs-McCarthy (1987) describes the difference between the approach of Natural Morphology and his own as that the former views one-to-one mapping as an ideal that languages strive towards, whereas for him, it serves as a starting point of linguistic description. Nevertheless, in both cases, this ideal concept stems from linguists’ description of morphology, not from empirical research on language learnability. Even Kusters (2003) notes that “[t]he ‘simplest imaginable relationship’, as Carstairs-McCarthy calls it, must be understood from a linguist’s perspective which does not necessarily imply that it is simple for all kinds of language users alike” (p.26). This emphasises that transparency is in the first place a theoretical descriptive of language, not a summary of learner behaviour. Don (2017) similarly notes that transparency is often attributed a notion of naturalness, optimality, simplicity or purity, but his stance is slightly different in that he is arguing against the claim that “simplex one-to-one correspondences between form and meaning are part of the core of the linguistic system” (p.133). Assuming that language is, in a way, ‘by default’

transparent or supposed to be transparent is in his view mistaken in light of the vast amount of non-transparency found in natural language systems. He allows for the possibility that there may be a learning preference for transparent structures, but that this preference is not emanating from the grammar itself. What the above remarks illustrate, however, is that we may need to make a distinction between the point of view of linguistic description, which sees transparent form-meaning mappings as ideal, and a learner's perspective, who may not find all kinds of non-transparency difficult to learn (Kusters, 2003). Since across the series of experiments presented in this thesis, transparency alone does not lead to better learnability, it appears that the strength of the common claims surrounding transparency stems from the side of linguistic description.

One of the main conclusions of this thesis is that transparency does not exist in isolation in natural languages, and that approaches which test its interplay with other factors in language acquisition appear to be a more promising avenue for future research. Goldschneider & DeKeyser (2001) already noted that trying to ascribe the observed morpheme acquisition order to a single factor was not fruitful in the 1970s and 1990s, and their approach reflects the assumption that a combination of factors better explains the phenomenon at hand. This is by no means a new suggestion. Other authors have made very similar statements, including Aksu-Koç & Slobin (1985) who note that correlated factors impacting on acquisition are a likely scenario, but cannot be studied in isolation when investigating a single language. Krashen et al. (1975) similarly suggest that different factors might interact with one another. With respect to the influence of word length on learning, Laufer (1990) remarks that

“[i]n a learning situation, however, it is hard to attribute the difficulty of learning a particular word to its length rather than to a variety of factors. It may also be that length becomes significant beyond a certain point, but it remains to be found out what point exactly” (p.298).

Finally, Gass & Selinker (2008) summarise factors that have been proposed to impact morpheme acquisition such as semantic complexity, frequency and perceptual salience and conclude that “[w]hat is more realistic, yet subject to empirical verification, is that these factors all contribute to acquisition order. What is then left to be determined is the relative weighting each factor has” (p.135).

These suggestions reflect what the results of the present study indicate as well: learning difficulty may result from many factors and the size of the paradigm may play

a role in determining when certain factors take effect. The findings of this thesis support the idea that the learnability of morphological systems stems from the combined effects of multiple predictors (Andersen, 1978; Kwon, 2005; Penke, 2012). The beneficial effect of factors like frequency and regularity on language acquisition has been demonstrated separately before (Culbertson & Newport, 2015; Ellis & Collins, 2009; Fehér et al., 2016; Hudson Kam & Newport, 2009; Larsen-Freeman, 2010; Pfänder et al., 2013; Saldana et al., 2021; Xanthos et. al., 2011). Additionally, a multitude of studies have been conducted that demonstrate the emergence of compositionality under learnability pressures, where the impact of a particular structure on learnability is more direct (Beckner et al., 2017; Chaabouni et al., 2020; Kirby et al., 2008; Kirby et al., 2015; Vogt, 2005). The findings from this thesis showing that a transparency bias might exist, but may be weak, support the view that transparency is not the single cause of learning ease.

## **6.7 Why paradigm size matters**

Having described difficulties with demonstrating an impact of transparency alone and the presence of multiple factors in natural language paradigms, it is easy to see why the size of the paradigm becomes important: it allows for additional factors to come into play, such as generalisability, frequency and regularity. In contrast to transparency, differences with respect to these factors increase as paradigms become larger: compositional systems will express meanings through the concatenation of a smaller inventory of distinct morphemes, while holistic systems will require more distinct morphemes (Brighton, 2002). The latter type of system is typically also more prone to affix variation, which will increase the number of different morphemes further as the paradigm grows (Plungian, 2008; Haspelmath, 2009). Relatedly, there will exist more regularity in a compositional than a holistic system.

The size of the paradigms tested in this thesis needed to be small in order to avoid the additional confound of differing numbers of morphemes to be learned. For example, three binary features would require the acquisition of 6 agglutinating, but 8 fusional morphemes, and more morphemes to be learned would likely contribute additional difficulty to the fusional system, obscuring the impact of transparency. For the same reason, it is not overly surprising that small inflectional systems do not differ

significantly in learnability. With regards to the size of CASE systems found in a sample of 47 natural languages, Plank (1986) has shown that small CASE paradigms occur similarly often in the form of agglutinating and fusional systems. However, the largest systems (11 CASE terms or more) in his sample all display agglutinating-type structures. What this suggests is that small systems are equally learnable in either morphological technique, but large systems are only learnable when they are of the agglutinating type. Consequently, an impact on learnability is to be expected, but it seems to be more prominent the larger the systems become, at a point where we are no longer dealing with transparency differences alone.

A study by Laaha & Gillis (2007) described in **Chapter 2** supports this finding, where they measure the number of categories in a sample of 9 languages. The strongly fusional and the strongly agglutinating languages express more form-type categories on nouns and verbs than the weakly fusional languages, but there is also a difference in the inflectionally rich languages. While languages in the strongly fusional and the strongly agglutinating groups mark a similar number of grammatical categories on nouns, the strongly agglutinating group marks about twice as many categories on verbs (between 19 and 37 categories depending on the language) compared to the strongly fusional group (between 15 and 18 depending on the language). This is in line with Plank's (1986) finding that the largest number of categories in a paradigm tends to be expressed by separative exponence, while smaller numbers of categories can be expressed well by both types of inflection. Kirby et al. (2015) similarly speculate that an increase in the meaning space in their iterated learning studies would exercise a stronger pressure for compressibility of the system. Overall, these different findings support the possibility that when the category system is small, as in the present experiments, both types of exponence are similarly common – and learnable.

Considering the possibility that multiple factors together make compositional systems easier to learn than fusional systems is in perfect accord with the claim that agglutinating structures (which display compositional structure) are learned better than fusional structures (Hengeveld & Leufkens, 2018; Igartua, 2015). This is the case because agglutination and fusion each comprise a multitude of characteristics, as outlined in **Section 1.2** and **Chapter 2**, hence allowing for these additional factors to impact on acquisition ease. The reason why the agglutinating condition in the experiments presented here did not yield higher accuracy rates than the fusional

condition likely is that the systems were too small for the benefits of frequency, generalisability and regularity, in combination with others like transparency, to take effect. In natural languages, however, we may well expect agglutinating morphology to be learned faster than fusional morphology. The evidence for this is rather compelling: consider for example the time difference in children acquiring Turkish versus Dutch verbal morphology (Aksu-Koç and Slobin, 1985; De Houwer & Gillis, 1998; Hengeveld & Leufkens, 2018) or the emergence of agglutinating structures in contact situations (Igartua, 2015). Compositionality, which underlies agglutination, has been shown to arise under pressures for learnability in a multitude of iterated learning studies and generalisability has been shown to be beneficial for learning (Beckner et al., 2017; Chaabouni et al., 2020; Kirby et al., 2008; Kirby et al., 2015; Vogt, 2005).

On a related note, it is then true after all that transparent structures are easier to learn than non-transparent structures, since agglutinating systems are typically transparent. However, they are not easier to learn for reasons of transparency. Put differently, transparency appears to be a feature of these systems, but not the root cause of acquisition ease. To avoid confusion, this is not what studies by Bittner et al. (2003), Korecky-Kröll & Dressler (2009), Igartua (2015), Narasimhan & Gullberg (2011) and others had in mind, though – their claim was that transparency does play an important role. However, what the findings from this thesis suggest is that the actual cause of learning ease needs to be searched for elsewhere. In this vein, transparency appears to be rather a symptom, a by-product, rather than the main cause of language change. Stating that languages become more transparent is not the same as saying that languages strive for transparency. Similarly, the findings presented here seem to suggest that learners find transparent structures easier to learn, but do not find transparency easier than non-transparency, at least not to a considerable extent. All is not lost for transparency, however. The experiments presented in this thesis consistently show a slight bias for transparency when the error patterns are concerned. Kempe & Brooks (2008) also show that transparent morphemes can provide cues to other related features and highlight systematicity in the morphological system to a learner. This suggests that transparency does play a role for acquisition, but that it is limited and likely interacts with other characteristics of morphological systems, paving the way for a variety of follow-up research.

## 6.8 Summary and outlook

Returning to the initial claims motivating this research, agglutinating morphology may indeed be easier to learn than fusional morphology for a variety of reasons outlined above. However, the results from five large experiments with over 600 participants in total presented here suggest that the learnability advantage does not stem from transparency, at least not exclusively, and that it is not necessarily dependent on transparent morphology in the L1. In the earlier studies presented as exemplary cases, transparent forms are always part of a larger inflectional paradigm, and often, other characteristics of these markers within the paradigm or the language more broadly have not been described in great detail. One may ask the question why the claim that transparent morphemes are easier to learn is so widespread, given that there appears to be little evidence for its immediate impact on learnability. To some extent, the popularity of the transparency claim may derive from the fact that linguistic theories take up a central part of linguistic research (Eddington, 2004). However, the experimental results presented in this thesis show that linguistic description does not automatically coincide with the behaviour of learners. Existing research also still makes it difficult to understand the nature of the potential correlation of transparency with factors such as generalisability and frequency, which have individually been shown to support learning as well.

Goldschneider & DeKeyser's (2001) research is an attempt to understand the interplay of several factors that likely play an important role for morpheme acquisition. However, while they are able to show that a combination of factors does have explanatory power, they are not able to test the contribution of each individual factors due to correlations in their model. Narasimhan & Gullberg (2011) pit frequency and transparency against each other, to investigate which factor learners may be more biased towards. This approach appears very relevant, but more data on the exact frequencies of the markers and their role in the language more broadly, such as their productivity and the predictability of the contexts they can occur in, would be insightful to consider as well. Korecky-Kröll & Dressler (2009) also distinguish different characteristics of individual plural markers, and research might benefit from studies with a larger subject pool to reduce the impact of idiosyncratic learner patterns. Nevertheless, naturalistic data makes it difficult to understand the exact extent to which a single characteristic of an inflectional marker impacts on learnability.

On a different note, there has been a certain trend to view language as a complex dynamic system, where different parts of the system interact and where particular characteristics of individual learners can offer insights into how well aggregate measures describe the learning reality (De Bot et al., 2007; Larsen-Freeman & Cameron, 2008). Without advocating for a particular approach, the underlying assumption that single factors cannot explain a complex picture appears adequate. Transparency is still being suggested as a powerful explanatory tool for learning ease at present, but there is little available direct evidence of the suggested causal relationship. It may be time for research in the area of L2 acquisition of inflectional morphology to cast a wider net and to test the effects of transparency more rigorously on the one hand, as well as describe its limitations on the other hand. The present thesis does not find an effect of transparency on learning overall, and only some evidence for a bias that may be weak, but it has only tested native speakers of three languages and it has used a specific inflectional paradigm.

Moreover, this study tested the end-product of acquiring morphology, measured in accuracy rates, not the types of processing strategies employed by the different learner groups. Testing acquisition of a small paradigm by participants at an early stage of learning, I do not find any differences in acquisition success between conditions. However, it is possible that processing strategies differed between participant groups even in this particular learning situation. Beyond this, the theoretical questions regarding the influence of L1 morphology on the processing of L2 morphology are nevertheless of scientific interest, especially because little is known about second language processing at low proficiency levels. Future research could complement this picture by contrasting different constellations of L1 and L2 morphology in beginning learners, to better understand when L1 morphology starts to play a role for L2 online and offline processing.

To conclude, transparency may play a role during acquisition, but we may want to ask when exactly and in what dimension. Experimental studies can tease apart these individual effects more easily and allow for more controlled tests. My research suggests that learners acquire transparent mappings between meanings and signals well, but equally well as non-transparent mappings when the meaning space is small. It would be interesting to be able to test larger meaning spaces and observe to what degree transparency in combination with other factors can explain learner behaviour. The paradigm used in the series of experiments presented in this thesis can be

adapted to include differences in the frequency or the generalisability of forms, for instance, and learnability differences with different paradigm sizes can be tested as well. Transparency on the system-level also leads to regularity, and in reverse, the effects of phenomena like stem and affix alternation or syncretism on learning should therefore also be considered. Thus, the effect size of different predictors could be tested in isolation, but also an interplay between multiple factors can be probed. The role of transparency itself also needs further investigation, to understand how much of the results presented here stem from the experimental design and how much they can be generalised to other contexts. Since some bias for transparency seems to exist, more research is needed to understand its scope. To put it in Bybee's (1985) words, "no explanation for linguistic phenomena is complete until a causal relation can be shown to exist between the principle proposed as explanation and the linguistic phenomena to be explained" (p.207). This thesis has provided suggestive evidence that the search for a link between transparency and learnability needs to continue in order to complement the decades of earlier studies on the acquisition of transparent and non-transparent morphology.

## Chapter 7

### Conclusion

That transparency has an impact on the acquisition of grammatical morphemes has been considered for many decades in language acquisition research, but transparency has rarely been probed under controlled conditions. A precise and comprehensive testing of the size of the effect of transparency on learnability was the central goal of this thesis. In order to illustrate the interplay of morpheme transparency in L1 and L2, this thesis began with an overview of the emergence of morphological typology, which aims to classify languages based on their morphological make-up, in **Chapter 1**. This has highlighted that morphological systems can differ considerably between languages, but that they are also varied language-internally. What this means for this thesis is that we will understand more from investigating subparts of the language, such as nominal morphology, or individual grammatical features, than considering whole languages. **Chapters 1** and **2** have also shown that morphemes in natural languages have multiple characteristics besides transparency, such as frequency and regularity, they occur in words of differing length and they can occur in paradigms with fewer or more morphemes. Against this backdrop, it was shown why it is important to assess the impact of transparency in a controlled artificial language learning paradigm: Earlier studies on first and second language acquisition mainly focused on natural languages, such as English in the morpheme order studies (Andersen, 1978; Brown, 1976; de Villiers & de Villiers, 1973; Larsen-Freeman, 1976), Dutch and Tamil (Narasimhan & Gullberg, 2011), Austrian German (Korecky-Kröll & Dressler, 2009) or a sample of typologically different languages (Laaha & Gillis, 2007). Suggestions about the impact of transparency on language change have also focused on samples of natural languages (Anttila, 1989; Igartua, 2015), as illustrated in **Section 2.1.5**.

The difficulty with the context of these claims is that morphemes in natural languages exhibit many characteristics besides transparency, which have individually been shown to impact learning ease, such as compositionality, generalisability and frequency. This was discussed in **Sections 2.2** and **2.3**. Similarly, research has found that characteristics of the L1, but especially also morphological characteristics of the L1, can influence acquisition of L2 morphology, as shown in **Sections 2.5** and **4.1.2**. In order to test whether experience with inflectional morphology in general, or with transparent, agglutinating morphology in particular, is necessary for a learner to be

able to profit from the transparency advantage, experiments were conducted with three groups of native speakers to represent different morphological techniques in the L1: English (isolating), Turkish (agglutinating) and Spanish (fusional). However, as **Chapters 3-5** illustrated, across all participant groups, no effect of transparency on overall learning was found and only a slight bias for transparent structures was revealed in the analysis of participants' errors. Importantly, however, this weak bias was as consistent as the lack of an overall effect, since it was also found in each separate analysis. In order to test whether potential difficulties of agglutinating systems adversely affected learning, two additional experiments were run which manipulated the length of the stimuli words and highlighted morpheme boundaries, since agglutinating words are typically longer and contain more morphemes which need to be segmented. However, neither of these two characteristics was found to impact acquisition of the agglutinating system.

Since the experimental design was carefully planned to reduce the influence of confounds like regularity as well as number and frequency of morphemes, along with avoiding the learning of novel phonotactics and word stems, the fact that transparency was repeatedly shown to be weak at best, with over 600 participants, raises serious questions regarding the common claim of a facilitating effect of transparency. As **Chapter 6** has illustrated, studies on the effect of transparency in natural languages did not account for the impact of confounds as precisely as the experiments in this thesis. It has also been shown that the claim itself may to a certain extent stem from linguistic theory, without having been put to test in a rigorous enough manner. For this reason, the present study presents a first step in the direction of more direct probing of transparency. Artificial language learning experiments with different miniature languages and other participant groups appear to be a fruitful avenue for future research, as well as testing the interplay of transparency and its confounds in these controlled settings. The nature of morphological processing across proficiency levels and different languages would contribute additional insights to this complex picture. The role of transparency in the acquisition of morphology has been investigated and asserted for many decades, but as the present study shows, there is more to be learned.

## 8. References

- Acarlar, F., & Johnston, J. R. (2011). Acquisition of Turkish grammatical morphology by children with developmental disorders: *Atypical acquisition of Turkish morphology*. *International Journal of Language & Communication Disorders*, 46(6), 728–738. <https://doi.org/10.1111/j.1460-6984.2011.00035.x>
- Aksu-Koç, A., & Slobin, D. (1985). The acquisition of Turkish. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 1, pp. 839–878). Hillsdale, NJ: Lawrence Erlbaum.
- Andersen, R. W. (1978). AN IMPLICATIONAL MODEL FOR SECOND LANGUAGE RESEARCH. *Language Learning*, 28(2), 221–282. <https://doi.org/10.1111/j.1467-1770.1978.tb00134.x>
- Anderson, S. R. (2015). The Morpheme: Its Nature and Use. In M. Baerman (Ed.), *The Oxford Handbook of Inflection* (pp. 10–34). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199591428.013.2>
- Anttila, R. (1989). *Historical and Comparative Linguistics*. John Benjamins Publishing Company.
- Atkinson, M., Smith, K., & Kirby, S. (2018). Adult Learning and Language Simplification. *Cognitive Science*, 42(8), 2818–2854. <https://doi.org/10.1111/cogs.12686>
- Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and Plurals in Dutch: Evidence for a Parallel Dual-Route Model. *Journal of Memory and Language*, 37(1), 94–117. <https://doi.org/10.1006/jmla.1997.2509>
- Baayen, R. H. & Linke, M. (2020). *An introduction to the generalized additive model*. Department of Linguistics, University of Tübingen. <http://www.sfs.uni-tuebingen.de/~hbaayen/publications/BaayenLinke2020.pdf>
- Bates D., Mächler M., Bolker B., & Walker S. (2015). “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01
- Beckner, C., Pierrehumbert, J. B., & Hay, J. (2017). The emergence of linguistic structure in an online iterated learning task. *Journal of Language Evolution*, 2(2), 160–176. <https://doi.org/10.1093/jole/lzx001>
- Bickel, B., & Nichols, J. (2009). Inflectional morphology. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 169–240). Cambridge University Press. DOI: 10.1017/CBO9780511618437.003.
- Bickel, B., & Nichols, J. (2013a). Fusion of Selected Inflectional Formatives. In: Dryer, Matthew S. & Haspelmath, Martin (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/20>, Accessed on 2021-09-30.)
- Bickel, B., & Nichols, J. (2013b). Exponence of Selected Inflectional Formatives. In: Dryer, Matthew S. & Haspelmath, Martin (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/21>, Accessed on 2021-04-29.)
- Bickel, B., & Nichols, J. (2013c). Inflectional Synthesis of the Verb. In: Dryer, Matthew S. & Haspelmath, Martin (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/22>, Accessed on 2021-09-30.)

- Bittner, D., Dressler, W., & Kilani-Schoch, M. (2003). *Development of verb inflection in first language acquisition. A cross-linguistic perspective*. DeGruyter.
- Blanco, N. J., & Sloutsky, V. M. (2019). Adaptive flexibility in category learning? Young children exhibit smaller costs of selective attention than adults. *Developmental Psychology*, 55(10), 2060–2076. <https://doi.org/10.1037/dev0000777>
- Blumenthal, A. & Kortmann, B. (2013). Die Verschiedenheit der Sprachen. In: Auer, P. (Ed.) *Sprachwissenschaft: Grammatik - Interaktion - Kognition*. Stuttgart: Metzler. 285-317.
- Blumenthal, A., Stojanoski, B., Martin, C. B., Cusack, R., & Köhler, S. (2018). Animacy and real-world size shape object representations in the human medial temporal lobes. *Human Brain Mapping*, 39(9), 3779–3792. <https://doi.org/10.1002/hbm.24212>
- Bonin, P., Gelin, M., Dioux, V., & Méot, A. (2019). “It is alive!” Evidence for animacy effects in semantic categorization and lexical decision. *Applied Psycholinguistics*, 40(4), 965–985. <https://doi.org/10.1017/S0142716419000092>
- Bosch, S., Krause, H., & Leminen, A. (2017). The time-course of morphosyntactic and semantic priming in late bilinguals: A study of German adjectives. *Bilingualism: Language and Cognition*, 20(3), 435–456. <https://doi.org/10.1017/S1366728916000055>
- Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language*, 29(5), 591–610. [https://doi.org/10.1016/0749-596X\(90\)90054-4](https://doi.org/10.1016/0749-596X(90)90054-4)
- Brighton, H. (2002). Compositional Syntax From Cultural Transmission. *Artificial Life*, 8(1), 25–54. <https://doi.org/10.1162/106454602753694756>
- Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, 2(3), 177–226. <https://doi.org/10.1016/j.plrev.2005.06.001>
- Brooks, P. J., Braine, M. D. S., Catalano, L., Brody, R. E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language*, 32(1), 76–95.
- Brown, R. (1973). Development of the First Language in the Human Species. *The American Psychologist*, 28(2), 97-106.
- Brown, R. (1976). *First Language: The Early Stages*. Harvard University Press.
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. J. Benjamins.
- Calvillo, D. P., & Jackson, R. E. (2014). Animacy, perceptual load, and inattentional blindness. *Psychonomic Bulletin & Review*, 21(3), 670–675. <https://doi.org/10.3758/s13423-013-0543-8>
- Carstairs-McCarthy, A. (1987). Allomorphy in inflexion. Croom Helm.
- Carstairs-McCarthy, A. (1992). *Current morphology*. Routledge.
- Chaabouni, R., Kharitonov, E., Bouchacourt, D., Dupoux, E., & Baroni, M. (2020). Compositionality and Generalization in Emergent Languages. *ArXiv:2004.09124 [Cs]*. <http://arxiv.org/abs/2004.09124>

- Chamorro, G., Sturt, P., & Sorace, A. (2016). Selectivity in L1 Attrition: Differential Object Marking in Spanish Near-Native Speakers of English. *Journal of Psycholinguistic Research*, 45(3), 697–715. <https://doi.org/10.1007/s10936-015-9372-4>
- Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological Structure in Native and Nonnative Language Processing. *Language Learning*, 60(1), 21–43. <https://doi.org/10.1111/j.1467-9922.2009.00550.x>
- Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some Input on the Easy/Difficult Grammar Question: An Empirical Study. *The Modern Language Journal*, 93(3), 336–353. <https://doi.org/10.1111/j.1540-4781.2009.00894.x>
- Comrie, B. (1981). *Language Universals and Linguistic Typology*. Basil Blackwell Publisher Limited.
- Corbett, G. G. (2013a). Sex-based and Non-sex-based Gender Systems. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/31>, Accessed on 2022-05-25.)
- Corbett, G. G. (2013b). Systems of Gender Assignment. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/32>, Accessed on 2022-05-28.)
- Croft, W. (2003). *Typology and Universals*. Cambridge University Press.
- Culbertson, J. (2012). Typological Universals as Reflections of Biased Learning: Evidence from Artificial Language Learning: Typological Universals as Reflections of Biased Learning. *Language and Linguistics Compass*, 6(5), 310–329. <https://doi.org/10.1002/Inc3.338>
- Culbertson, J., & Newport, E. L. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, 139, 71–82. <https://doi.org/10.1016/j.cognition.2015.02.007>
- Culbertson, J., & Schuler, K. (2019). Artificial Language Learning in Children. *Annual Review of Linguistics*, 5(1), 353–373. <https://doi.org/10.1146/annurev-linguistics-011718-012329>
- De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(01), 7. <https://doi.org/10.1017/S1366728906002732>
- de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, 2(3), 267–278. <https://doi.org/10.1007/BF01067106>
- De Diego-Balaguer, R., & Rodriguez-Fornells, A. (2010). Contributions to the Functional Neuroanatomy of Morphosyntactic Processing in L2. *Language Learning*, 60(1), 231–259. <https://doi.org/10.1111/j.1467-9922.2009.00557.x>
- De Houwer, A. & Gillis, S. (1998). Dutch child language: An overview. In Gillis, S. & De Houwer, A. (Eds.), *The acquisition of Dutch*, 1–100. Amsterdam: John Benjamins.
- DeKeyser, R. M. (2005). What Makes Learning Second-Language Grammar Difficult? A Review of Issues: *Language Learning*, 55(S1), 1–25. <https://doi.org/10.1111/j.0023-8333.2005.00294.x>
- Don, J. (2017). What causes languages to be transparent? *Language Sciences*, 60, 133–143. <https://doi.org/10.1016/j.langsci.2017.01.004>

- Dressler, W. U. (2005). Morphological Typology and First Language Acquisition: Some Mutual Challenges. *On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4)*. URL: <http://morbo.lingue.unibo.it/mmm/>
- Dressler, W. U. (2007). Introduction. In Laaha, S. & Gillis, S. (Eds.) *Typological perspectives on the acquisition of noun and verb morphology*. Antwerp Papers in Linguistics.
- Dressler, W. U. (2012). On the acquisition of inflectional morphology: Introduction. *Morphology*, 22(1), 1–8. <https://doi.org/10.1007/s11525-011-9198-1>
- Dressler, W. U., Mayerthaler, W., Panagl, O., & Wurzel, U. (1987). *Leitmotifs in Natural Morphology*. Cambridge University Press.
- Dressler, W. U., Kilani-Schoch, M., Pestal, L., & Pöchtrager, M. (2006). *On the Typology of Inflection Class Systems*. *Folia Linguistica*. Mouton de Gruyter.
- Dulay, H. C., & Burt, M. K. (1974). NATURAL SEQUENCES IN CHILD SECOND LANGUAGE ACQUISITION <sup>1</sup>. *Language Learning*, 24(1), 37–53. <https://doi.org/10.1111/j.1467-1770.1974.tb00234.x>
- Eddington, D. (2004). *Spanish Phonology and Morphology: Experimental and quantitative perspectives*. John Benjamins Publishing Company.
- Ellis, N., & Collins, L. (2009). Input and Second Language Acquisition: The Roles of Frequency, Form, and Function Introduction to the Special Issue. *The Modern Language Journal*, 93(3), 329–335. <https://doi.org/10.1111/j.1540-4781.2009.00893.x>
- Etymonline.com. (n.d.). Agglutination. In *Online Etymological Dictionary*. Retrieved May 20, 2022, from <https://www.etymonline.com/word/agglutination>.
- Fábregas, A. (2018). Word Phenomena: Category Definition and Word Formation. In K. L. Geeslin (Ed.), *The Cambridge Handbook of Spanish Linguistics* (1st ed., pp. 261–285). Cambridge University Press. <https://doi.org/10.1017/9781316779194.013>
- Fehér, O., Wonnacott, E., & Smith, K. (2016). Structural priming in artificial languages and the regularisation of unpredictable variation. *Journal of Memory and Language*, 91, 158–180. <https://doi.org/10.1016/j.jml.2016.06.002>
- Fortescue, M., Mithun, M., & Evans, N. (2017). *Introduction* (M. Fortescue, M. Mithun, & N. Evans, Eds.; Vol. 1). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199683208.013.1>
- Fruchter, J., Stockall, L., & Marantz, A. (2013). MEG masked priming evidence for form-based decomposition of irregular verbs. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00798>
- Fruchter, J., & Marantz, A. (2015). Decomposition, lookup, and recombination: MEG evidence for the Full Decomposition model of complex visual word recognition. *Brain and Language*, 143, 81–96. <https://doi.org/10.1016/j.bandl.2015.03.001>
- Garcia Garcia, M. (2018). Nominal and verbal parameters in the diachrony of differential object marking in Spanish. In Serzant, I. A. & Witzlack-Makarevich, A. (Eds.). *Diachrony of Differential Argument Marking* (pp.209-242). Language Science Press.
- Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed). Routledge/Taylor and Francis Group.
- Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the “Natural Order of L2 Morpheme Acquisition” in English: A Meta-analysis of Multiple Determinants. *Language Learning*, 51(1), 50.

- Gor, K., & Cook, S. (2010). Nonnative Processing of Verbal Morphology: In Search of Regularity. *Language Learning*, 60(1), 88–126.  
<https://doi.org/10.1111/j.1467-9922.2009.00552.x>
- Gor, K., & Jackson, S. (2013). Morphological decomposition and lexical access in a native and second language: A nesting doll effect. *Language and Cognitive Processes*, 28(7), 1065–1091. <https://doi.org/10.1080/01690965.2013.776696>
- Greenberg, J. H. (1960). A Quantitative Approach to the Morphological Typology of Language. *International Journal of American Linguistics*, 26(3), 178-194.
- Harley, H., & Ritter, E. (2002). Person and Number in Pronouns: A Feature-Geometric Analysis. *Language*, 78(3), 482–526.  
<https://doi.org/10.1353/lan.2002.0158>
- Harris, Z. S. (1951). *Structural linguistics*. The University of Chicago Press.
- Haspelmath, M. (2009). An Empirical Test of the Agglutination Hypothesis. In S. Scalise, E. Magni, & A. Bisetto (Eds.), *Universals of Language Today* (Vol. 76, pp. 13–29). Springer Netherlands.  
[https://doi.org/10.1007/978-1-4020-8825-4\\_2](https://doi.org/10.1007/978-1-4020-8825-4_2)
- Haspelmath, M. (2013). Occurrence of Nominal Plurality. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/34>, Accessed on 2022-05-25.)
- Haspelmath, M., König, E., Oesterreicher, W., & Raible, W. (Eds.). (2008). *Language Typology and Language Universals*. De Gruyter.
- Haspelmath, M., & Michaelis, S. M. (2017). Analytic and synthetic: Typological change in varieties of European languages. In I. Buchstaller & B. Siebenhaar (Eds.), *Studies in Language Variation* (Vol. 19, pp. 3–22). John Benjamins Publishing Company. <https://doi.org/10.1075/silv.19.01has>
- Havas, V., Waris, O., Vaquero, L., Rodríguez-Fornells, A., & Laine, M. (2015). Morphological learning in a novel language: A cross-language comparison. *Quarterly Journal of Experimental Psychology*, 68(7), 1426–1441.  
<https://doi.org/10.1080/17470218.2014.983531>
- Hengeveld, K., & Leufkens, S. (2018). Transparent and non-transparent languages. *Folia Linguistica*, 52(1), 139–175. <https://doi.org/10.1515/flin-2018-0003>
- Hualde, J. I. & Ortiz de Urbina, J. (Ed.). (2003). *A grammar of Basque*. Mouton de Gruyter.
- Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30–66.  
<https://doi.org/10.1016/j.cogpsych.2009.01.001>
- Igartua, I. (2015). From cumulative to separative exponence in inflection: Reversing the morphological cycle. *Language*, 91(3), 676–722.  
<https://doi.org/10.1353/lan.2015.0032>
- Iggesen, O. A. (2013). Number of Cases. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info/chapter/49>, Accessed on 2022-05-31.)
- Jacob, G., Şafak, D. F., Demir, O., & Kirkıcı, B. (2019). Preserved morphological processing in heritage speakers: A masked priming study on Turkish. *Second Language Research*, 35(2), 173–194.  
<https://doi.org/10.1177/0267658318764535>
- Janssen, B. E. (2016). *The acquisition of gender and case in Polish and Russian: A study of monolingual and bilingual children*. Uitgeverij Pegasus.

- Jarvis, S. (2015). 2. The Scope of Transfer Research. In L. Yu & T. Odlin (Eds.), *New Perspectives on Transfer in Second Language Learning* (pp. 17–48). Multilingual Matters. <https://doi.org/10.21832/9781783094349-004>
- Johnson, T., Gao, K., Smith, K., Rabagliati, H., & Culbertson, J. (2020). *Predictive structure or paradigm size? Investigating the effects of i-complexity and e-complexity on the learnability of morphological systems* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/qns6v>
- Kaivapalu, A., & Martin, M. (2007). Morphology in transition: Plural inflection of Finnish nouns by Estonian and Russian learners. *Acta Linguistica Hungarica*, 54(2), 129–156. <https://doi.org/10.1556/ALing.54.2007.2.2>
- Karlsson, F., Miestamo, M. & Sinnemäki, K. (2008). Introduction: The problem of language complexity. In Miestamo, M., Sinnemäki, K. & Karlsson, F. (Eds.) *Language Complexity. Typology, contact, change*. John Benjamins.
- Kempe, V., & Brooks, P. J. (2008). Second Language Learning of Complex Inflectional Systems. *Language Learning*, 58(4), 703–746. <https://doi.org/10.1111/j.1467-9922.2008.00477.x>
- Kimppa, L., Shtyrov, Y., Hut, S. C. A., Hedlund, L., Leminen, M., & Leminen, A. (2019). Acquisition of L2 morphology by adult language learners. *Cortex*, 116, 74–90. <https://doi.org/10.1016/j.cortex.2019.01.012>
- Kirby, S. (2000). Syntax Without Natural Selection: How Compositionality Emerges from Vocabulary in a Population of Learners. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *The Evolutionary Emergence of Language* (1st ed., pp. 303–323). Cambridge University Press. <https://doi.org/10.1017/CBO9780511606441.019>
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686. <https://doi.org/10.1073/pnas.0707835105>
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102. <https://doi.org/10.1016/j.cognition.2015.03.016>
- Korecky-Kröll, K., & Dressler, W. U. (2009). The acquisition of number and case in Austrian German nouns. In U. Stephany & M. D. Voeikova (Eds.), *Development of Nominal Inflection in First Language Acquisition* (Vol. 30, pp. 265–302). Mouton de Gruyter. <https://doi.org/10.1515/9783110217117.265>
- Kortmann, B. (2005). *English Linguistics: Essentials. Anglistik, Amerikanistik*. Cornelsen.
- Krashen, S. D., Madden, C., & Bailey, N. (1975). Theoretical aspects of grammatical sequencing. In M. K. Burt & H. C. Dulay (Eds.), *On TESOL '75: New directions in second language learning, teaching and bilingual education. Selected Papers from the Annual TESOL Convention (9th, Los Angeles, CA, March 4-9, 1975)* (pp.44-54). Teachers of English to Speakers of Other Languages.
- Krause, E., & von Heusinger, K. (2019). Gradient Effects of Animacy on Differential Object Marking in Turkish. *Open Linguistics*, 5(1), 171–190. <https://doi.org/10.1515/opli-2019-0011>
- Kusters, C. W. (2003). Linguistic complexity: The influence of social change on verbal inflection. LOT.

- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60(6):1126–1141.
- Krifka M. (2001). Compositionality. In: Wilson RA, Keil F, editors. *The MIT encyclopaedia of the cognitive sciences*. Cambridge, MA: MIT Press.
- Kwon, E. (2010). The “Natural Order” of Morpheme Acquisition: A Historical Survey and Discussion of Three Putative Determinants. *Working Papers in TESOL and Applied Linguistics*, 36(2), 179–204.  
<https://doi.org/10.21559/AELLK.2010.36.2.011>
- Laaha, S., & Gillis, S. (Eds.) (2007). *Typological perspectives on the acquisition of noun and verb morphology*. Antwerp Papers in Linguistics.
- Larsen-Freeman, D. E. (1976). AN EXPLANATION FOR THE MORPHEME ACQUISITION ORDER OF SECOND LANGUAGE LEARNERS. *Language Learning*, 26(1), 125–134. <https://doi.org/10.1111/j.1467-1770.1976.tb00264.x>
- Larsen-Freeman, D. (2002). MAKING SENSE OF FREQUENCY. *Studies in Second Language Acquisition*, 24(2), 275–285.  
<https://doi.org/10.1017/S0272263102002127>
- Larsen-Freeman, D. (2010). Not So Fast: A Discussion of L2 Morpheme Processing and Acquisition. *Language Learning*, 60(1), 221–230.  
<https://doi.org/10.1111/j.1467-9922.2009.00556.x>
- Larsen-Freeman, D., & Cameron, L. (2008). Research Methodology on Language Development from a Complex Systems Perspective. *The Modern Language Journal*, 92(2), 200–213. <https://doi.org/10.1111/j.1540-4781.2008.00714.x>
- Laufer, B. (1990). Some Intralexical Factors that Effect the Learning of Words. *International Review of Applied Linguistics in Language Teaching*, 28(4).
- Lehtonen, M., & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 6(3), 213–225. <https://doi.org/10.1017/S1366728903001147>
- Lehtonen, M., Niska, H., Wande, E., Niemi, J., & Laine, M. (2006). Recognition of Inflected Words in a Morphologically Limited Language: Frequency Effects in Monolinguals and Bilinguals. *Journal of Psycholinguistic Research*, 35(2), 121–146. <https://doi.org/10.1007/s10936-005-9008-1>
- Leonard, L. B. (2014). Specific Language Impairment Across Languages. *Child Development Perspectives*, 8(1), 1–5. <https://doi.org/10.1111/cdep.12053>
- Leung, J. H. C., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition*, 33(1), 33–55. <https://doi.org/10.1017/S0272263110000525>
- Leung, J. H. C., & Williams, J. N. (2012). Constraints on Implicit Learning of Grammatical Form-Meaning Connections: Implicit Learning of Grammatical Form-Meaning Connections. *Language Learning*, 62(2), 634–662.  
<https://doi.org/10.1111/j.1467-9922.2011.00637.x>
- Levshina, N. (2019). *Linguistic Frankenstein, or How to test universal constraints without real languages*. Zenodo. <https://doi.org/10.5281/ZENODO.2583820>
- Lewis, G., Solomyak, O., & Marantz, A. (2011). The neural basis of obligatory decomposition of suffixed words. *Brain and Language*, 118(3), 118–127.  
<https://doi.org/10.1016/j.bandl.2011.04.004>
- Liang, L., & Chen, B. (2014). Processing morphologically complex words in second-language learners: The effect of proficiency. *Acta Psychologica*, 150, 69–79.  
<https://doi.org/10.1016/j.actpsy.2014.04.009>

- Litman, L., Moss, A., Rosenzweig, C., & Robinson, J. (2021). Reply to MTurk, Prolific or panels? Choosing the right audience for online research. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3775075>
- Lowie, W. (2000). Cross-linguistic influence on morphology in the bilingual mental lexicon. *Studia Linguistica*, 54(2), 175–185. <https://doi.org/10.1111/1467-9582.00058>
- Luk, Z. P., & Shirai, Y. (2009). Is the Acquisition Order of Grammatical Morphemes Impervious to L1 Knowledge? Evidence From the Acquisition of Plural -s, Articles, and Possessive 's. *Language Learning*, 34.
- Lupyan, G., & Dale, R. (2010). Language Structure Is Partly Determined by Social Structure. *PLoS ONE*, 5(1), e8559. <https://doi.org/10.1371/journal.pone.0008559>
- Marsden, E., Williams, J., & Liu, X. (2013). LEARNING NOVEL MORPHOLOGY: The Role of meaning and orientation of attention at initial exposure. *Studies in Second Language Acquisition*, 35(4), 619–654. <https://doi.org/10.1017/S0272263113000296>
- McWhorter, J. (2007). *Language Interrupted*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195309805.001.0001>
- Mel'čuk, I. (2013). The notion of inflection and the expression of nominal gender in Spanish. *Studies in Language*, 37(4), 736–763. <https://doi.org/10.1075/sl.37.4.02mel>
- Miestamo, M., Sinnemäki, K., & Karlsson, F. (Eds.). (2008). *Language complexity: Typology, contact, change*. John Benjamins Pub. Co.
- Moreno-Sandoval, A. & Goñi-Menoyo, J. M. (2002). Spanish Inflectional Morphology in DATR. *Journal of Logic, Language, and Information*, 11(1), 79-105.
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. CRC Press/Taylor & Francis Group.
- Moreton, E., & Pater, J. (2012a). Structure and Substance in Artificial-phonology Learning, Part I: Structure: Structure and Substance in Artificial-Phonology Learning, Part I. *Language and Linguistics Compass*, 6(11), 686–701. <https://doi.org/10.1002/lnc3.363>
- Moreton, E., & Pater, J. (2012b). Structure and Substance in Artificial-Phonology Learning, Part II: Substance: Structure and Substance in Artificial-Phonology Learning, Part II. *Language and Linguistics Compass*, 6(11), 702–718. <https://doi.org/10.1002/lnc3.366>
- Narasimhan, B., & Gullberg, M. (2011). The role of input frequency and semantic transparency in the acquisition of verb meaning: Evidence from placement verbs in Tamil and Dutch. *Journal of Child Language*, 38(3), 504–532. <https://doi.org/10.1017/S0305000910000164>
- Neophytou, K., Manouilidou, C., Stockall, L., & Marantz, A. (2018). Syntactic and semantic restrictions on morphological recomposition: MEG evidence from Greek. *Brain and Language*, 183, 11–20. <https://doi.org/10.1016/j.bandl.2018.05.003>
- Peer, E., Rothschild, D., Evernden, Z., Gordon, A., & Damer, E. (n.d.). *Data Quality of Platforms and Panels for Online Behavioral Research*. 46.
- Penke, M. (2012). *The Acquisition of Inflectional Morphology*. <https://doi.org/10.13140/RG.2.2.21194.06085>
- Pfänder, S., Behrens, H., Auer, P., Jacob, D., Kailuweit, R., Konieczny, L., Kortmann, B., Mair, C., & Strube, G. (2013). Erfahrung zählt. Frequenzeffekte

- in der Sprache – ein Werkstattbericht. *Zeitschrift für Literaturwissenschaft und Linguistik*, 43(1), 7–32. <https://doi.org/10.1007/BF03379870>
- Plank, F. (1983). Transparent versus functional encoding of grammatical relations. *Linguistische Berichte* (86).
- Plank, F. (1986). PARADIGM SIZE, MORPHOLOGICAL TYPOLOGY, AND UNIVERSAL ECONOMY. *San Diego*, 20.
- Plank, F. (1999). *Split morphology: How agglutination and flexion mix*. *Linguistic Typology*, 3, 279-340.
- Plungian, V. A. (2008). Agglutination and flection. In Haspelmath, M., Koenig, E., Oesterreicher, W. (Eds). *Language Typology and Universals*. (pp.669-678). DeGruyter. <https://doi.org/10.1515/9783110194036-012>
- Portin, M., Lehtonen, M., & Laine, M. (n.d.). Processing of inflected nouns in late bilinguals. *Applied Psycholinguistics*, 23.
- Portin, M., Lehtonen, M., Harrer, G., Wande, E., Niemi, J., & Laine, M. (2008). L1 effects on the processing of inflected nouns in L2. *Acta Psychologica*, 128(3), 452–465. <https://doi.org/10.1016/j.actpsy.2007.07.003>
- Prévost, P., & White, L. (2000). Missing Surface Inflection or Impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16(2), 103–133. <https://doi.org/10.1191/026765800677556046>
- R Core Team (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. *Cognition*, 182, 151–164. <https://doi.org/10.1016/j.cognition.2018.09.010>
- Rebuschat, P., & Williams, J. N. (2012). Implicit Learning in Second Language Acquisition. In C. A. Chappelle (Ed.), *The Encyclopedia of Applied Linguistics* (p. wbeal0529). Blackwell Publishing Ltd. <https://doi.org/10.1002/9781405198431.wbeal0529>
- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2017). Distributional learning of subcategories in an artificial grammar: Category generalization and subcategory restrictions. *Journal of Memory and Language*, 97, 17–29. <https://doi.org/10.1016/j.jml.2017.07.006>
- Revithiadou, A., Spyropoulos, V., & Markopoulos, G. (2017). From Fusion to Agglutination: The Case of Asia Minor Greek. *Transactions of the Philological Society*, 115(3), 297–335. <https://doi.org/10.1111/1467-968X.12091>
- Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 22(1), 1–29. <https://doi.org/10.1191/0267658306sr259oa>
- Saldana, C., Smith, K., Kirby, S., & Culbertson, J. (2021). Is Regularization Uniform across Linguistic Levels? Comparing Learning and Production of Unconditioned Probabilistic Variation in Morphology and Word Order. *Language Learning and Development*, 17(2), 158–188. <https://doi.org/10.1080/15475441.2021.1876697>
- Sapir, E. (1921). *Language: An Introduction to the Study of Speech*.
- Seyfarth, S., Ackerman, F., & Malouf, R. (2014). Implicative organization facilitates morphological learning. *Annual Meeting of the Berkeley Linguistics Society*, 40, 480. <https://doi.org/10.3765/bls.v40i0.3154>

- Sims-Williams, H. (2016). Analogical Levelling and Optimisation: The Treatment of Pointless Lexical Allomorphy in Greek. *Transactions of the Philological Society*, 114(3), 315–338. <https://doi.org/10.1111/1467-968X.12078>
- Sims-Williams, H., & Enger, H.-O. (2021). The loss of inflection as grammar complication: Evidence from Mainland Scandinavian. *Diachronica*, 38(1), 111–150. <https://doi.org/10.1075/dia.19050.sim>
- Stewart, T. (2018). Structuralism. In J. Audring & F. Masini (Eds.), *The Oxford Handbook of Morphological Theory* (pp. 84–104). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199668984.013.11>
- Stoll, S. (2015). Inflectional Morphology in Language Acquisition. In M. Baerman (Ed.), *The Oxford Handbook of Inflection* (pp. 350–374). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199591428.013.15>
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11(6), 1090–1098. <https://doi.org/10.3758/BF03196742>
- Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23(7–8), 942–971. <https://doi.org/10.1080/01690960802069730>
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.) *Studies of Child Language Development*. Holt, Rinehart & Winston.
- Slobin, D. I. (1980). The repeated path between transparency and opacity in language. In Bellugi, U. & Studdert-Kennedy (Eds.). *Signed and spoken language: biological constraints on linguistic form*, Verlag Chemie GmbH, 229-243.
- Smith, K. (2018). The cognitive prerequisites for language: Insights from iterated learning. *Current Opinion in Behavioral Sciences*, 21, 154–160. <https://doi.org/10.1016/j.cobeha.2018.05.003>
- Soto, F. A., Wasserman, E. A. (2010). Missing the forest for the trees: Object discrimination learning blocks categorization learning. *Psychological Science*, 21(10), 1510–1517.
- Strickland, B. (2017). Language reflects “core” cognition: A new theory about the origin of cross-linguistic regularities. *Cognitive science*, 41(1):70–101.
- Szabo, Z. G. (2021). *COMPOSITIONALITY AS SUPERVENIENCE*. 32.
- Szagun, G. (2001). Learning different regularities: The acquisition of noun plurals by German-speaking children. *First Language*, 21(62), 109–141. <https://doi.org/10.1177/014272370102106201>
- van Schaik, G. (2020). *The Oxford Turkish Grammar* (1st ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780198851509.001.0001>
- Vihman, V.-A., & Nelson, D. (2019). Effects of Animacy in Grammar and Cognition: Introduction to Special Issue. *Open Linguistics*, 5(1), 260–267. <https://doi.org/10.1515/opli-2019-0015>
- Vincent, N. (2010). Compositionality and change. In *The Routledge Handbook of Historical Linguistics*. Routledge. <https://doi.org/10.4324/9781315794013.ch3>
- Vogt, P. (2005). On the Acquisition and Evolution of Compositional Languages: Sparse Input and the Productive Creativity of Children. *Adaptive Behavior*, 13(4), 325–346. <https://doi.org/10.1177/105971230501300403>
- Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1

- and L2 speakers of English. *Journal of Phonetics*, 70, 86–116.  
<https://doi.org/10.1016/j.wocn.2018.03.002>
- Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., Gagarina, N., Hrzica, G., Ketrez, F. N., Kilani-Schoch, M., Korecky-Kröll, K., Kovačević, M., Laalo, K., Palmović, M., Pfeiler, B., Voeikova, M. D., & Dressler, W. U. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Language*, 31(4), 461–479.  
<https://doi.org/10.1177/0142723711409976>
- Zweig, E., & Pykkänen, L. (2009). A visual M170 effect of morphological complexity. *Language and Cognitive Processes*, 24(3), 412–439.  
<https://doi.org/10.1080/01690960802180420>

## 9. Appendix

### Study information form used with native speakers of English:



Linguistics & English Language  
University of Edinburgh

---

#### Consent for Participation in Experiments, Data Use, and Data Storage

---

Study title:	Learning part of a new language
Principal Investigator:	Dr. Jennifer Culbertson
Researcher collecting current data:	Svenja Wagner

**Purpose and nature of the study.** You are about to participate in a study that looks at how people learn and understand new languages. It involves looking at images, listening to audio, and clicking on words. Once you finish, you will be asked a few questions about yourself (e.g. age, gender, language background) and the experience you had during the study. Your responses will be recorded electronically. The session should take about 30 minutes. You will be given full instructions shortly.

**Compensation.** You will be paid \$4 for your participation in this study.

**Risks and benefits.** There are no known risks to participation in this study other than those encountered in everyday life. Other than the payment mentioned, the only benefits to you personally are those you draw from making a contribution to our knowledge about language and its use.

**Confidentiality.** The data we collect will not be associated with your name or with any other personal details that might identify you. All data will be anonymized. We will temporarily collect your worker ID in order to prevent repeat participation, however we will not permanently store this ID.

**Voluntary participation and right to withdraw.** Your participation is voluntary, and you may withdraw from the study at any time and for any reason. If you withdraw from the study before data collection is complete (and do not submit the HIT) we are not able to pay you. If you choose to withdraw afterwards, please use the contact information below. We will delete your data and there will be no penalty or loss of benefits to which you are otherwise entitled.

**Contact information.** This research is being conducted by the above-listed researchers at the University of Edinburgh. The researchers can be contacted at [s1581727@sms.ed.ac.uk](mailto:s1581727@sms.ed.ac.uk) for questions or to report a research-related problem. Contact the Linguistics & English Language Ethics committee at 0131 651 5510 or [lcl.ethics@ed.ac.uk](mailto:lcl.ethics@ed.ac.uk) if you have concerns regarding your rights as a participant in the research.

If you have any questions about what you've just read, please contact the researchers before continuing.

Thank you for your help!

By accepting this HIT, you consent to the following:

1. I agree that the **anonymized data** I produce may be **kept permanently in Edinburgh University archives** and used **for the specific research project** for which they are collected.
2. I agree that the **anonymized data** I produce may be **used by the above-named researchers, as well as by other qualified researchers, for teaching or research purposes, and in professional presentations and publications.**
3. I agree that that the **anonymized data** I produce **may be made publicly available for general use**, e.g. used in radio or television broadcasts, or put on the world-wide web.
4. I understand that I have the **right to terminate this session** at any point. If I choose to **withdraw after completing the study**, my data will be deleted at that time.

## Turkish native speaker tests:

### 1. Aşağıdakilerden hangi kelimenin yazılışı doğrudur?

- A. Bu masada tertemiz oldu.
- B. İş arkadaşına karşı kindar bir tutum takındı.
- C. Mine'ye görmeyeli yıllar oldu.
- D. Bir çift söz etmeme izin varmı?

### 2. Aşağıdaki cümlelerden hangisinde yazım hatası bulunmamaktadır?

- A. Kitabım sen de kalsın.
- B. Çantanıda almalısın.
- C. Evimiz'de kedi besliyoruz.
- D. Bahçede oynadık.

### 3. Aşağıdaki seçeneklerden hangisi doğrudur?

- A. Akıl zihinde deęi baştır.
- B. Akıl yaşıta deęil baştır.
- C. Akıl başta deęil yaşıtır.
- D. Akıl başta deęil kafadadır.

### 4. Aşağıdakilerden hangi kelimenin yazılışı doğrudur?

- A. Melike senin arkadaş mı?
- B. Senin yaşınız nedir?
- C. Bende dün baş ağrımısı vardı.
- D. Çiçekleri masanın üzerine koydum.

## Spanish native speaker tests:

### 1. Por favor, haga "clic" en la frase que es correcta en Español.

- A. Si estaba viviendo en Alaska, estaría usando botas de invierno por ocho meses porque el clima es muy frío.
- B. Si estuviera viviendo en Alaska, estaría usando botas de invierno por ocho meses porque la clima es muy frío.
- C. Si estuviera viviendo en Alaska, estaría usando botas de invierno por ocho meses porque el clima es muy frío.
- D. Si estuviera viviendo en Alaska, estaré usando botas de invierno por ocho meses porque el clima es muy frío.

### 2. Por favor, haga "clic" en la frase que es correcta en Español.

- A. Cuando viví en mi departamento pasado, una vez necesité ir al doctor porque me lastimé la mano en el gimnasio, pero era domingo y terminé yendo al hospital.
- B. Cuando vivía en mi departamento pasado, una vez necesité ir al doctor porque me lastimé el mano en el gimnasio, pero era domingo y terminé yendo al hospital.
- C. Cuando solía vivía en mi departamento pasado, una vez necesité ir al doctor porque me lastimé la mano en el gimnasio, pero era domingo y terminé yendo al hospital.
- D. Cuando vivía en mi departamento pasado, una vez necesité ir al doctor porque me lastimé la mano en el gimnasio, pero era domingo y terminé yendo al hospital.

### 3. Por favor, haga "clic" en la frase que es correcta en Español.

- A. El punto principal que ella está tratando de hacer con los ejemplos que tomó de los libros de sus colegas de la oficina son que los efectos del tratamiento alternativo son cuestionados.
- B. El punto principal que ella está tratando de hacer con los ejemplos que tomó de los libros de sus colegas de la oficina es que los efectos del tratamiento alternativo son cuestionados.
- C. El punto principal que ella está tratando de hacer con los ejemplos que tomaba de los libros de sus colegas de la oficina es que los efectos del tratamiento alternativo son cuestionados.
- D. El punto principal que ella está tratando de hacer con los ejemplos que tomó de los libros de sus colegas de la oficina es que los efectos del tratamiento alternativo son cuestionados.

### 4. ¿En qué situación diría alguien esta frase?

**"No se hizo la miel para la boca del asno."**

- A. Cuando tu mejor amiga quiere invitar a su novio a cenar en un lugar especial, pero tú sabes que su novio no lo apreciará porque no la trata bien.
- B. Cuando has invertido mucho tiempo trabajando en un proyecto y finalmente recibes una recompensa por tu trabajo.
- C. Cuando uno de tus mejores amigos comparte comida contigo, simplemente porque está a punto de caducar.

D. Cuando a los niños pequeños se les dice que no coman miel porque les da dolor de estómago.